



UNIVERSIDADE FEDERAL DO ESTADO DO RIO DE JANEIRO – UNIRIO

Centro de Ciências Humanas e Sociais - CCH

Programa de Pós-Graduação em Memória Social - PPGMS

Camila Guimarães Dantas

**Criptografias da memória:
um estudo teórico-prático sobre o arquivamento da web no Brasil**

Rio de Janeiro

2014

Camila Guimarães Dantas

**Criptografias da memória:
um estudo teórico-prático sobre o arquivamento da web no Brasil**

Tese apresentada e ao Programa de
Pós Graduação em Memória Social da
Universidade Federal do Estado do Rio de
Janeiro como requisito parcial para conclusão
do doutorado em Memória Social.

Orientadora: professora Dr^a Vera Dodebei

Rio de Janeiro

2014

D192 Dantas, Camila Guimarães.
Criptografias da memória : um estudo teórico-prático sobre o arquivamento da web no Brasil / Camila Guimarães Dantas, 2014.
228 f. ; 30 cm

Orientadora: Vera Dodebei.
Tese (Doutorado em Memória Social) – Universidade Federal do Estado do Rio de Janeiro, Rio de Janeiro, 2015.

1. Internet - Brasil - História. 2. Ferramentas de busca na Web.
3. Arquivo. 4. Memória digital. I. Dodebei, Vera. II. Universidade Federal do Estado do de Janeiro. Centro de Ciências Humanas e Sociais. Programa de Pós-Graduação em Memória Social. III. Título.

CDD – 005.430981

Camila Guimarães Dantas

Criptografias da memória: um estudo teórico-prático sobre o arquivamento da web no Brasil

Rio de Janeiro _____ de _____ de _____

Prof^a Dr^a Vera Dodebei

orientadora

UNIRIO

Prof^a Dr^a Regina Abreu

UNIRIO

Prof^a Dr^a Leila Beatriz Ribeiro

UNIRIO

Prof^a Dr^a Marieta de Moraes Ferreira

titular

CPDOC/FGV

Prof^a Dr^a Luciana Quillet Heymann

CPDOC/FGV

Para Vilma e Ronaldo.

```
<!DOCTYPE html><html xmlns="http://www.w3.org/1999/xhtml"><head>meta content="
Archive-It -
Brazilian Web Engines (1997-2013)
</title><link type="text/css" rel="stylesheet" href="/static/css/glob
//
var con // AGRADecIMENTOS
var origURI "/collections/4266";
var origQuery = "";
var origQueryTerms = "";
function init()
{
    agora.init({url: origURI, origQuery, origQueryTerms,
        "#search-textbox", "#curr-tab", "#all-search-results", 500});
}
function autoSelect()
{
    Vilma Guimarães, Ronaldo Dantas, Lúcia Pontes,
    Gabriel Lima, Luciana Guimarães, Jorge van Kats,
    Lucas&Iara&Laiz, Raquel Temporal, Toinho Castro,
    Luciane Lessa, Val Medeiros, Família van Kats ();
}
$(init);
</script></head><body><div class="container_l2 logo-wrapper"><h1 class="log
Archive-It
Navigation</a><
This web collection preserves captures from Brazilian web engines. The snapshot
</p><p><b>Subject: </b><a href="/explore?show=Collection&fc=meta Sul
Arts & Humanities</a>, <a href="/explore?show=Collection&fc=met
Society & Culture</a>, <a href="/explore?show=Collection&fc=met
// Computers & Technology</a>, <a href="/explore?show=Collection&f
Web History, Search Engines, Brazil</a></p><div style="display: inh
({ Renato Franco, Francisco Ramos de Farias, Diana de Souza
Pinto, Jô Gondar, Wolfgang Bock, José Bessa, Anneke Smelik,
Camila G. Dantas</a></p></div></div><div id="all-search
Liedeke Plaate, Ivan da Costa Marques
//
});
if (agora) {
    agora.setState({
        "Adriene Baron Tacla, Aline Marinho, Guiselle Martha Starink,
        Débora Terra, Patrícia Pechini, Inti Maya"
        page : 1,
        sort : "f_url.asc",
        //
        tab : "Sites",
        rootUrl : "",
        {waybackAccessPoint : "4266"
    });
    Ivana Arandjelovic, Alessandra Carvalho,
    Adriana Trindade, Chintia Araujo, Armelle Enders, Anita
    Luchesi, Caroline van de Steeg, Roberta Rego, Mariana Torres,
    Karina Galindo, Juliana Almeida
    <a title="http://busca.uol.com.br/" href="https://wayback.archive-it.org/42
    Loading Wayback Capture Info...
    </p>};<div style="display: inherit;" class="moreMetadata"></div></div><div
    Title:
    cadê?
    </h3><h3 class="url">URL:
    <a title="http://cade.com.br/" href="https://wayback.archive-it.org/4266/*/
    It was created in 1995 by Gustavo Viberte and Fabio de Oliveira. in Ri
```

<http://www.everythingyouseeinthepast.com>

Rafaël Rozendaal, 2003



RESUMO

Este trabalho investiga as práticas contemporâneas implicadas no arquivamento, na preservação e no acesso ao passado da web. A noção de criptografias da memória é tomada como uma via de análise para pensar as operações mnemônicas em suporte digital. Os enredamentos entre os aspectos técnicos e culturais que atravessam as telas do presente, ao mesmo tempo em que condicionam os modos de arquivamento e acesso ao passado recente, são tomados nesta pesquisa como um ponto crucial na composição desta memória criptografada. Neste trabalho procurou-se chamar atenção para as intrincadas configurações entre os aspectos técnicos (os formatos e padrões) e culturais (como uma sociedade define o que é memória digital). No caso do Brasil identificamos uma confluência entre a noção de memória digital e acervo digitalizando do que decorrem consequências de ordem prática, como a inexistência de coleções de páginas da web em instituições nacionais. É uma investigação teórico-prática, na medida em que a reflexão levou à realização de uma coleção digital de páginas das ferramentas de busca do Brasil, entre 1997 e 2013. A coleção Buscas.br foi criada através de uma parceria com o Internet Archive, que possibilitou a execução da curadoria e disponibilizou a plataforma Archive-It para tornar os dados acessíveis através da pesquisa por palavras. A realização desta coleção, que pode ser pesquisada por qualquer pessoa interessada na história da web no Brasil, propiciou um contato com a materialidade das fontes digitais, alimentando nossa reflexão teórica. A partir dessa experimentação, e também do levantamento bibliográfico, defendemos que as criptografias da memória no que se refere ao arquivamento da web são constituídas de elementos de caráter menos plástico ou dinâmico do que costuma ser descrito na literatura sobre a memória digital.

Palavras-chave: Memória digital, internet, arquivo, coleção, ferramentas de busca.

ABSTRACT

This work investigates the contemporary practices involved in archiving, preserving, and accessing the web's history. The notion of encrypted memory is taken as a pathway to think about the operations of memory in digital materiality, seeking to draw attention on the entanglements between the technical and cultural aspects that run through the screens. It is a theoretical and practical research because the reflection led to the realization of a digital collection of websites from Brazilian search tools from 1997 to 2013. The collection (Brazilian Web Engines or Buscas .br, in Portuguese) was created through a partnership with the Internet Archive, which enabled the implementation; and Archive -It has provided the platform to make the data accessible through word search. The process of building this collection, which can be searched by anyone interested in the history of the web in Brazil, provided a contact with the materiality of digital sources and it has brought new input to our theoretical analysis. We argue that the encrypted memory present in this collection has not a plastic or dynamic quality as it is often assumed by the literature related to digital memory.

Key words: digital memory, internet, web archiving, collection, search engines.

LISTA DE IMAGENS

Imagem 1 - Página inicial do site Europeia em 2013	26
Imagem 2 - Herança cultural para futuras gerações	29
Imagem 3 - Distribuição dos 43 membros do IIPC	36
Imagem 4 - Página inicial do website <i>Wayback Machine</i>	40
Imagem 5 - Área de pesquisa do site Arquivo da Web Portuguesa	42
Imagem 6 - Não há registro de publicidade nos sites arquivados no Arquivo da Web Portuguesa	43
Imagem 7 - Página do Fórum de Cultura Digital, acessada através do Internet Archive	49
Imagem 8 - Página de apresentação do grupo de Memória Digital	50
Imagem 9 - Fragmento da postagem de Leo Germani no grupo de Memória Digital ...	53
Imagem 10 - Fragmento de página do Fórum com a nossa primeira intervenção	55
Imagem 11 - Página inicial do website <i>Memoryshare</i>	66
Imagem 12 - Selecionado um registro no website <i>Memoryshare</i>	67

Imagem 13 - Acesso à memória “batom, pó e palmada” no website <i>Memoryshare</i> . . .	68
Imagem 14 - Imagem e conteúdo de um postal publicado no website <i>PostSecret</i>	70
Imagem 15 - Página do website <i>Internet Archive</i>	78
Imagem 16 - Arquivo do website <i>Post Secret</i>	78
Imagem 17 - Registro do arquivo do website <i>Post Secret</i> no website <i>Internet Archive</i>	79
Imagem 18 - Resultado de busca no website <i>UK Web Archive</i>	80
Imagem 19 - Registro do arquivo do website <i>WW2 People’s War</i> no website <i>UK Web Archive</i>	80
Imagem 20 - Resultado da busca pelo website <i>WW2 People’s War</i> no website <i>Wayback Machine</i>	81
Imagem 21 - Gráfico relacionando usuários das web 1.0 e 2.0 à quantidade de usuários do Facebook	89
Imagem 22 - Esquema afirmando as continuidades entre web 1.0 e web 2.0	91
Imagem 23 - Resultado da busca pelo site <i>archive.org</i> nos registros do website <i>Wayback Machine</i>	95
Imagem 24 - Registro de 1997 do website <i>Internet Archive</i> no próprio website	97

Imagem 25 - Registro de 2000 do website <i>Internet Archive</i> no próprio website	99
Imagem 26 - Diagrama: a arquitetura para registro e acesso às coleções do website <i>Internet Archive</i>	101
Imagem 27 - Esquema de hierarquia técnica na nova estética do arquivo proposto por Wolfgang Ernest.	105
Imagem 28 - Arquivo sobre as eleição presidencial no Brasil do website <i>Library of Congress</i>	112
Imagem 29 - Coleção <i>Brazilian Web Engines (1997-2013)</i>	121
Imagem 30 - Resultado da pesquisa por <i>História do Brasil</i> no website <i>Archive-It</i> ...	122
Imagem 31 - Tela do <i>Archive-It</i> com a listagem de seeds a partir dos quais pode-se consultar todas as páginas arquivadas	124
Imagem 32 - Esquema: visualização da arquitetura da coleção na interface do website <i>Archive-It</i>	127
Imagem 33 - Tela inicial da categoria <i>História do Cadê?</i> no website <i>Archive-It</i>	128
Imagem 34 - Ruína da página <i>Anti-Explocação</i> no website <i>Internet Archive</i>	129
Imagem 35 - Informações do código da página <i>Anti-Explocação</i> em HTML no website <i>Internet Archive</i>	130
Imagem 36 - Enquete realizada pela revista <i>Internet.Br</i> apontava o <i>Cadê?</i> como a ferramenta de busca mais utilizadas pelos leitores do periódico	133

Imagem 37 - Algumas imagens da web brasileira em 1997 nos arquivos do website <i>Internet Archive</i>	134
Imagem 38 - Página inicial do Yahoo! em 1997.....	135
Imagem 39 - Página inicial do Cadê? em 1997.....	136
Imagem 40 - Catálogo de 1997 traz referência ao <i>Cadê?</i>	139
Imagem 41 - Incoerência temporal no registro do website <i>Cadê?</i> no <i>Internet Archive</i>	142
Imagem 42 - Falha no arquivamento do website <i>Radix</i>	143
Imagem 43 - Trecho do manifesto Glitch, de Menckman (2010).....	145
Imagem 44 - Imagem glitch da coleção <i>Buscas.Br</i>	146

LISTA DE TABELAS

Tabela 1 - Instituições com projetos específicos de arquivamentos da web.	31
Tabela 2 - As quatro principais ferramentas de busca	115
Tabela 3 - Diferentes características do processo de armazenamento e acesso aos dados dos websites <i>Internet Archive</i> e <i>Archive-It</i>	118

SUMÁRIO

Introdução	16
1. Arquivos digitais:	
práticas emergentes e conceitos em elaboração	21
1.1 Registros digitais –	
as configurações possíveis do legado digitalizado e nascido digital	21
1.1.1 Iniciativas relevantes de digitalização:	
Europeana e Biblioteca Digital Mundial	25
1.1.2 Experiências de arquivamento da Web.....	30
1.1.3 A Biblioteca Real da Holanda e o Arquivo Nacional Português	38
1.2 Discursos de patrimonialização e novos conceitos no Brasil.....	44
1.2.1 Noções de memória digital no Brasil a partir	
do estudo de caso do Fórum da Cultura Digital	44
1.2.2 O silêncio sobre o passado da web no debate relacionado	
à memória digital no Brasil	58
2. Memória digital: um campo de múltiplas conexões	60
2.1 Enredamentos teóricos no campo de estudo das memórias sociais	61
2.2 Memórias anônimas na web.....	64
2.2.1 <i>Memoryshare</i> ou é possível compartilhar memórias via internet?.....	65
2.2.2 <i>PostSecret</i> - o que são os segredos públicos da nossa rede digital?	70
2.2.3 Anonimato e registro: como pensar o coletivo?	72
2.3 O desgaste da documentação: revisitando o portal BBC People’s War	76
2.4 Os desgastes das memórias digitais	77

3. Páginas e protocolos enredados:	
a web e a constituição do seu maior arquivo.	83
3.1 Visões da rede: muito além do ciberespaço	83
3.1.1 A retórica da ruptura na cronologia da web	88
3.1.2 A ampliação da rede e os oráculos digitais	92
3.2 Enredos e dados do Internet Archive	94
3.2.1 O projeto do Internet Archive: legado cultural em dados	95
3.2.2 A máquina do tempo do Internet Archive	100
3.2.3 O curto-circuito da memória em dados digitais	103
3.2.4 O Internet Archive é um arquivo?	107
4. Buscas.Br: uma experimentação no arquivamento	
de páginas da web brasileira	109
4.1 Uma coleção de páginas?	111
4.1.1 Como fazer uma coleção de páginas?	
Relato de uma incursão no emaranhado de dígitos	115
4.2 A arquitetura da coleção	123
4.2.1 Um percurso na coleção	126
4.2.2 O caso de um catálogo brasileiro da web 1.0: o portal de buscas Cadê?	131
4.2.3 O Cadê? a partir do relato de um dos seus fundadores	132
4.3 Em busca das buscas	142
Considerações Finais	147
Anexos I	149
Referências	216

INTRODUÇÃO

A possibilidade de ter um endereço eletrônico (no caso, de conteúdo artístico) como epígrafe deste trabalho nos remete diretamente à nossa temática. O website, página ou ainda sítio como querem alguns em português, já se tornou um objeto cultural facilmente reconhecido pelos leitores do século XXI. Esta investigação teórico-prática se propõe a pensar os modos pelos quais este objeto vem sendo arquivado, tornando-se um elo importante nas intrincadas operações de produção da memória documental do passado recente.

Descrever um percurso de pesquisa é tentar ligar os pontos entre os interesses inexplicáveis por determinado objeto, as possibilidades teóricas e as limitações metodológicas. Ao ligar os pontos, constrói-se, palavra a palavra, dígito a dígito, uma possibilidade de abrir novos percursos. A minha busca nestas veredas digitais inicia-se com o meu projeto de mestrado sobre as configurações da memória social em registros digitais. A dissertação de mestrado foi uma oportunidade para mergulhar em alguns objetos que representavam possibilidades temáticas variadas como, por exemplo, os arquivos gerados pelo trauma e os usos das ferramentas digitais pela historiografia.

Ao elaborar o projeto de doutorado, uma preocupação era de enveredar nas possibilidades conceituais existentes para, a partir disto, pensar a relação da memória com as novas tecnologias da informação. Na pesquisa de mestrado enfrentei alguns desafios relacionados à organização da informação eletrônica e também ao acesso à documentação nascida digital. Nesta primeira experiência pude constatar que há uma espécie de “poeira” entre as páginas arquivadas em servidores e também entre as mais diversas telas consultadas ao redor do mundo. Há uma poeira digital com uma materialidade própria a este suporte. Refletir sobre o “gosto” desta poeira é caminhar na trilha lançada há mais de duas décadas por Arlette Farge (1989), quando investigou a pesquisa nos arquivos e chamou a atenção para uma fenomenologia do espaço e do contato com os documentos de arquivo. As páginas dilaceradas, as fraturas dos documentos e o ambiente solene e silencioso das salas de consulta do arquivo foram tomados pela historiadora francesa como elementos fundamentais para se compreender o espaço do arquivo. Essas reflexões foram uma inspiração na construção de uma problemática que interroga os modos de contato entre o pesquisador e a documentação digital.

E o que seriam as criptografias da memória? A criptografia, como uma arte de codificar textos, a princípio de caráter militar, tem suas origens na Antiguidade. No século XX a criptografia de dados digitais tornou-se uma prática corriqueira, regida por um sistema de algoritmos, para a troca segura de informações das mais diversas naturezas (pessoal, comercial, militar, etc). Neste trabalho propomos o termo “criptografias da memória” para pensar as práticas con-

temporâneas de registro do passado através das novas tecnologias da informação, mais especificamente da web. Isto não implica na pressuposição de que estaríamos na iminência de decodificar segredos ou códigos perdidos como em um romance policial. A criptografia funciona como uma metáfora para este mundo informacional do qual iremos nos aproximar. Na verdade, o que procuramos sugerir com o termo, e que deverá ficar explicitado ao final do estudo, é a necessidade de pensar as operações relacionadas à memória social em suporte digital a partir de uma encruzilhada que permita o diálogo entre a literatura que trata do tema da memória social, no campo das ciências humanas, e uma já significativa produção sobre as novas mídias digitais.

Tal diálogo será realizado a partir de um ponto preciso de intersecção de tais campos do saber: os arquivos constituídos por objetos nascidos digitais. A designação “nascido digital” aplica-se a arquivos que se originaram na rede mundial de computadores e que não possuem uma contrapartida física, embora alguns deles possuam uma estrutura material, necessária à manutenção dos dados e das interfaces próprias deste tipo de arquivo. Assim, esta pesquisa não se deterá em questões relacionadas diretamente à digitalização de acervos e às complexas operações que envolvem a transposição de documentos analógicos para o ambiente digital.

O interesse por questões metodológicas, assim como a disposição para observar as subjetividades implicadas no processo de construção do objeto, já estavam presentes desde o início do doutorado. Outras questões foram acrescentadas à medida que se procurava delinear melhor o objeto de pesquisa. Trabalhamos na direção de uma redução na escala, tal como sugerida por uma série de pesquisadores em contrapartida aos discursos generalistas e mistificadores sobre as novas tecnologias da informação. Encontrar o ponto de inserção no emaranhado de campos de saberes e possíveis objetos empíricos exigiu uma sondagem ampla do campo bibliográfico, bem como uma imersão em práticas diferenciadas de pesquisa, associadas principalmente ao campo das Humanidades Digitais.

O mergulho na literatura relacionada às novas mídias, em especial nos estudos de software (MANOVICH, 2001; 2008) e na história da Web e dos websites (BRUGGER, 2009; ROGERS, 2010), levou-me a buscar um objeto que desse a ver as especificidades técnicas mais gerais do campo da produção das memórias documentais em suporte digital. A importância do suporte no entendimento dos objetos culturais já vem sendo sublinhada por muitos ramos da pesquisa social. A abordagem da História Cultural, proposta por Roger Chartier (1990), postula justamente a necessidade de se pensar na materialidade dos objetos, sejam eles livros ou documentos eletrônicos.

A problemática da pesquisa começou a se delinear ao mesmo tempo em que os desafios para estudar as fontes ficavam cada vez mais nítidos. Numa primeira tentativa de mapeamento

das instituições de memória no Brasil, através da participação no grupo Memória Digital (2010) do Forum da Cultura Digital, não conseguimos identificar nenhum projeto em andamento que utilizasse conteúdos nascidos digitais. A questão que se colocava era então: como pesquisar o passado da web no Brasil, a partir de fontes nascidas digitais? A única via possível seria recorrer ao Internet Archive, único repositório de páginas brasileiras de acesso público às páginas da web 1.0. Deste ponto de partida surgiram duas ordens de inquietação. Em primeiro lugar, perguntava-me de que maneiras o Internet Archive constituía-se como um acervo/memória da web brasileira. Em segundo lugar, diante de milhões de páginas arquivadas, que ponto específico deveria ser interrogado neste passado recente encapsulado nos vestígios digitais?

Enfrentar essas inquietações significou desbravar o Internet Archive, descobrindo todo um território feito de ruínas e deparando-me com bugs, links quebrados e impossibilidades técnicas de lidar com documentos que pareciam abrir janelas para um certa performance do registro. Se o processo de flunar nas telas e visitar arquivos online foi, sem dúvida, fundamental, é preciso apontar também a importância da participação em workshops de pesquisa, que abriram novos horizontes ao trabalho. A experiência no *Digital Humanities Summer Institute* (Canadá) em 2010 e depois a participação em dois workshops, em 2011 e 2013, no *Digital Methods Initiative* (Universidade de Amsterdam), tornaram possível a experimentação com práticas de pesquisa coletivas onde os projetos eram pensados e postos em prática por uma equipe multidisciplinar que contava com pesquisadores de ciências humanas, programadores e designers.

A imersão neste contexto de pesquisa encorajou-me a desenvolver um projeto experimental com o objetivo de ter um contato mais próximo com os processos técnicos envolvidos na preservação e no acesso aos dados nascidos digitais. O que poderia ser pesquisado? A partir de um levantamento inicial percebemos a ausência de iniciativas de arquivamento da web no Brasil, o que instigou a nossa curiosidade nesta direção, delimitando um campo de interesse da pesquisa. Uma segunda temática viria balizar de modo importante o trabalho e definir o nosso recorte – a questão das ferramentas de busca. O motor desta segunda delimitação está relacionado ao lugar no qual nos inserimos, no sentido atribuído por Michel de Certeau; ao tempo presente, que nos impõe questionamentos. Hoje, ao teclar, pesquisar, ou simplesmente navegar na rede mundial de computadores vivenciamos um cotidiano marcado pela crescente importância das ferramentas de busca. Autores diversos (LOVINK, 2012; HALAVASIS, 2008; VAIDHYANATHANOU, 2011) são unânimes em apontar o papel central e centralizador por elas exercido na experiência atual da web, com destaque absoluto para a empresa que domina o setor – o Google.

A configuração contemporânea da web, marcada pela atividade das ferramentas de busca nos levou a interrogar o passado. Como seria entrar numa web ainda não conformada pelas fer-

ramentas de busca atuais? Quais eram as vias de entrada da web brasileira antes da Era Google? Para responder a estas perguntas é que nos propomos a criar uma coleção de páginas da web. A pesquisa bibliográfica e as entrevistas realizadas possibilitaram o recolhimento de informações sobre as principais ferramentas brasileiras, na época chamadas de engenhos ou mecanismos de busca. A evidente distinção entre os modos de funcionamento da web 1.0, com limitados recursos de armazenamento se compararmos com os dias de hoje, traz questões relevantes sobre a relação entre a memória e a web. Se hoje o Google enfrenta processos relacionados ao Direito ao Esquecimento, por manter uma base de dados incrivelmente atualizada, boa parte das páginas dos primeiros buscadores brasileiros desapareceram. Das quatro ferramentas escolhidas para compor nosso corpus – UOL, Radix, Cadê e Todo.br – apenas uma delas, o UOL, permanece ativa. O que restou dos demais não foram base de dados complexas, mas rastros digitais, referências escritas e uma memória oral.

Propomos nesta pesquisa, a partir de um estudo de caso, explorar este território das ferramentas de busca, como forma de investigar as criptografias da memória e contribuir para o campo interdisciplinar da memória social. Uma fonte de inspiração fundamental para este trabalho é a obra de Wolfgang Ernst (2005, 2013) que contém uma elaborada reflexão sobre os regimes de arquivamento na contemporaneidade. A perspectiva arqueológica de Ernst, segundo a qual a materialidade das mídias deve ser perscrutada, inspirou o impulso de colocar a mão na massa, ou melhor, nos dados. Além desta sinalização metodológica, os postulados apresentados por Ernst sobre uma nova estética do arquivo trouxeram questões importantes para o delineamento de nossa problemática.

O fazer da coleção nos colocou diante das especificidades técnicas do processo de arquivamento digital que, de outro modo, não poderíamos conhecer abrindo com isso uma série de novas reflexões. Ao construirmos a coleção Buscas.Br, ou *Brazilian Web Engines*, elaborada em parceria com o Archive-It, procuramos investigar de que modo aconteciam as dinâmicas entre as páginas arquivadas. Depois de um longo processo, descrito no quarto capítulo, a coleção finalmente está disponível em: <https://www.archive-it.org/collections/4266>. Mas não é o artefato técnico por si só que importa, e sim de que maneira ele nos permite enveredar nas operações contemporâneas implicadas nas relações entre o passado e o presente nas telas da web. A coleção nos colocou diante dos vestígios do recente passado das buscas na web 1.0 no Brasil. Ao mesmo tempo nos expôs um modo muito novo de preservação, cuja relativa precariedade e morosidade tornam visível uma face da web completamente oposta à rapidez e instantaneidade que lhe são usualmente atribuídas.

Organização dos capítulos

No primeiro capítulo – *Arquivos digitais: práticas emergentes e conceitos em elaboração* – realizamos um mapeamento de importantes experiências contemporâneas de arquivamento da web no mundo. O objetivo desta parte foi proporcionar um contexto mais amplo no qual insere-se este estudo. Ao mesmo tempo, como a intenção era a de agir no campo de pesquisa, interessou-nos investigar em maior profundidade o cenário nacional. A discussão sobre a emergência do conceito de patrimônio digital com conotações específicas no Brasil, relacionadas à digitalização dos acervos, é um ponto central deste capítulo.

No segundo capítulo – *Memória digital: um campo de múltiplas conexões* – exploramos os entrelaçamentos conceituais do campo interdisciplinar no qual se constitui este estudo, através da exploração de projetos de caráter memorial. Buscou-se, ao ampliar o panorama sobre a cultura de memória na qual estamos inseridos, elucidar a noção de criptografias da memória. Tal noção é trabalhada neste capítulo associada, sobretudo, ao conceito de tecnologias da memória proposto por Marita Sturken (1999).

Em *Páginas e protocolos enredados*: a web e a constituição do seu maior arquivo, terceiro capítulo, aprofundamos a discussão sobre as criptografias da memória a partir do estudo do Internet Archive e das práticas de preservação digital instituídas. Neste capítulo abordamos as práticas de preservação das páginas, enfocando os formatos e protocolos como elementos constitutivos de uma tecnologia da memória, a qual denominamos de criptografias da memória.

No quarto capítulo – *Buscas.Br: uma experimentação no arquivamento de páginas da web brasileira* – apresentamos a coleção *Buscas.Br*, resultado da vertente experimental da tese. Nesta etapa, partimos do relato da experiência de elaboração do arquivo para propor uma reflexão teórica sobre as especificidades do arquivamento digital. Traçamos um percurso na documentação nascida digital da coleção, como forma de dar a ver não somente a riqueza do material ali contido, mas igualmente suas lacunas, mostrando o caráter fraturado das memórias em código binário. A incorporação da qualidade do *Glitch* na noção de criptografias só foi possível a partir da experimentação realizada nesta coleção. Neste capítulo realizamos ainda uma breve exploração do conteúdo da coleção, com foco no caso da ferramenta Cadê?, buscando cruzar os registros digitais com a documentação oral e escrita pesquisadas.

O Anexo Digital procura apresentar visualmente alguns mementos do percurso de elaboração da coleção. Sua realização é também um esforço no sentido de apropriar-se das imagens da pesquisa, das subjetividades envolvidas e criar uma forma de expressar essa navegação experimental.

1. ARQUIVOS DIGITAIS: PRÁTICAS EMERGENTES E CONCEITOS EM ELABORAÇÃO

Neste primeiro capítulo nossa principal intenção é propor um mapeamento de determinados conceitos e práticas associados aos arquivos digitais. Ao mapear, estaremos, na verdade, selecionando os pontos de referência para pensar a nossa problemática. Como até agora vem sendo pensada a questão da preservação dos documentos eletrônicos? Quais as principais instituições envolvidas no debate, no mundo e no Brasil? Quais práticas estão sendo efetivamente perpetuadas neste campo emergente? Como situar tais práticas numa perspectiva histórica da própria rede mundial de computadores? Quais as principais visões que estão sendo elaboradas sobre o passado da rede?

No primeiro item, abordaremos a questão dos registros em bits e da emergência do conceito de patrimônio digital. O papel da Unesco e de seus documentos oficiais será analisado como forma de entendermos as políticas públicas sobre o tema. Ao mesmo tempo, pretende-se articular o debate atual com as discussões já em voga acerca da categoria do patrimônio.

No segundo item, a proposta é lançar um olhar sobre as práticas e conceitos ligados aos arquivos digitais que têm se constituído no Brasil. No decorrer da pesquisa encontramos muitas experiências relevantes de digitalização de acervos, porém não conseguimos encontrar, no âmbito nacional, um arquivo especificamente voltado para os objetos nascidos digitais. Esta é uma das questões que nortearam o nosso mapeamento do conceito de patrimônio digital no Brasil, realizado principalmente através do estudo de caso do Fórum da Cultura Digital Brasileira.

1.1 Registros digitais – as configurações possíveis do legado digitalizado e nascido digital

Embora o foco da nossa pesquisa seja os registros em código binário que circulam na Web, em especial aqueles que são nativos deste meio, tais como aqueles originados/relativos a *blogs*, *tweets*, mensagens eletrônicas, etc), trabalharemos neste capítulo com uma categoria mais ampla de “registros digitais”, uma vez que o nosso objetivo é mapear a construção da noção de patrimônio digital. Dentro do vasto universo dos registros digitais, podem, certamente, ser qualificados como patrimônio não apenas os objetos produzidos para a rede mundial de computadores, mas também aqueles que foram digitalizados, tais como manuscritos, e também

objetos multimídia, armazenados em suporte digital mais estável como o *CD-Rom*, que permite apenas a leitura dos dados. Iniciaremos procurando situar os debates atuais sobre o patrimônio digital no contexto mais ampliado da discussão em torno do patrimônio.

Os estudos sobre a categoria patrimônio têm indicado a sua invenção moderna, remontando historicamente a um amplo processo que se inicia no século XVIII e encontra na França do século XIX um vasto campo de desenvolvimento e institucionalização. A defesa moderna da preservação patrimonial se fez atrelada a uma “retórica da perda”, isto é, do alarme diante da possibilidade de desaparecimento de bens imóveis qualificados como de relevância histórica; o que levou a uma ação ativa por parte de organizações internacionais e nacionais (GONÇALVES, 2003:23). No mundo ocidental houve, desde o século XIX, uma expansão notória das preocupações patrimoniais, muitas vezes associadas a projetos nacionais ou identitários (CHOAY, 2001).

A partir do século XX, a UNESCO desempenhou um papel fundamental nas discussões e políticas ligadas ao patrimônio. A *Convenção sobre a Proteção do Patrimônio Cultural e Natural* de 1972 constitui um marco importante, ao estabelecer os parâmetros para uma definição de patrimônio – atribuído a bens naturais e culturais – dando início a uma Lista do Patrimônio Mundial e criando um Fundo do Patrimônio Mundial. É somente em 1989 que a herança cultural intangível (saberes, práticas e manifestações culturais diversas) passa a ser considerada, com a *Recomendação sobre a Salvaguarda da Cultura Tradicional e Popular*. A partir daí uma série de esforços são realizados para avançar nas discussões e ações até que em 2003, a UNESCO estabelece a *Convenção para a Salvaguarda do Patrimônio Cultural Imaterial*, que estabelece uma série de diretrizes para identificar e perpetuar o patrimônio imaterial através de diversas linhas de ação (propostas educativas, estímulo a projetos nacionais, cooperação internacional, etc.).

É também no ano de 2003 que a Unesco se pronuncia sobre o patrimônio digital, lançando a *Carta sobre a Preservação do Patrimônio Digital*. Trata-se de uma defesa bem aliçada da preservação do patrimônio digital, tendo como pressuposto a sua própria existência. A carta parte do reconhecimento de que na atualidade, muitos bens culturais “são cada vez mais produzidos, distribuídos, acessados e mantidos em forma digital, criando um novo legado – o patrimônio digital” (UNESCO, 2003). O documento inclui em sua definição do patrimônio digital, como se pode constatar no texto abaixo transcrito, tanto aquilo que é produzido originalmente com tecnologia digital como os materiais de diferentes mídias que foram convertidos para o formato digital.

Recursos de conhecimento ou expressão humana, seja cultural, educacional, científico e administrativo, ou abrangendo a informação técnica, legal, médica e outros tipos de informação, são cada vez mais criados digitalmente, ou convertidos de sua forma analógica original à forma digital. Quando os recursos são criados em forma digital não existe outro formato, além do digital original.

Materiais digitais incluem textos, bases de dados, imagens estáticas e com movimento, áudios, gráficos, *softwares* e páginas Web, entre uma ampla e crescente variedade de formatos. Eles geralmente são passageiros e requerem produção, manutenção e gerenciamento intencionais para serem preservados.

Muitos desses materiais são de valor e significância duradouros, e por isso constituem um patrimônio que deve ser protegido e preservado para as gerações atual e futura (UNESCO, 2003).

É importante situar o pronunciamento da UNESCO sobre o patrimônio digital em articulação a todo um contexto internacional no qual a preservação de arquivos digitais já contava com um conjunto de iniciativas em curso e constituía um campo de debates em pleno desenvolvimento/expansão. Experiências de arquivamento da internet tais como o *Internet Archive* e o *Pandora (Preserving and Accessing Networked Documentary Resources of Austrália)*, desenvolvido pela Biblioteca Nacional da Austrália, ambas iniciadas na década de 1990, estavam em estágio avançado. Também havia esforços no sentido de pensar políticas e práticas voltadas para o patrimônio digital, através de organizações como a *Digital Preservation Coalition (DPC)*, criada em 2001, no Reino Unido, com o objetivo de promover o intercâmbio entre diversas instituições e regulamentar esta área de atuação.

O documento da UNESCO chama atenção para dificuldades intrínsecas à tecnologia digital, como o problema da obsolescência de *softwares* e *hardwares*, que se opõe à intenção de preservação. Aponta também para a facilidade de alteração do documento em suporte digital, destacando a necessidade de se estabelecer práticas capazes de garantir a autenticidade dos mesmos. Os Estados-membros da UNESCO são convocados a desenvolver políticas voltadas para os seus patrimônios digitais e designar instituições para administrá-los. No Brasil, o Conselho Nacional de Arquivos também lançou um documento (*Carta para a Preservação do Patrimônio Arquivístico Digital – Preservar para Garantir o Acesso*) em defesa da preservação do patrimônio digital. O texto destaca igualmente os riscos de perda diante da veloz modificação de suportes tecnológicos e propõe linhas possíveis de ação para preservar tal patrimônio (CONARQ, 2004).

A carta da UNESCO coloca-se certamente como um posicionamento em prol do desenvolvimento de políticas internacionais que favoreçam a preservação do patrimônio digital. Mas a definição daquilo que em meio ao vasto universo das informações digitais efetivamente deve ser alvo de preservação se constitui – como acontece em relação a todo e qualquer patrimônio – a partir de um território/campo de negociações e disputas. Como afirma Dodebei, “a atribuição de valores documentais e patrimoniais aos objetos produzidos pela sociedade não pode ser pensada de modo definitivo, uma vez que a ação documental/ patrimonial é produto de relação circunstancial” (2011, s/p). Isto implica dizer que são as iniciativas de preservação proporcionadas a um dado objeto digital que irão lhe trazer um valor patrimonial, que será construído por várias instâncias da sociedade. O documento da UNESCO deixa transparecer esse terreno móvel e movediço da valoração do patrimônio em seu artigo 7 – *Definindo o que deve ser mantido* que, ao apresentar critérios tão amplos e vagos para a preservação, dá a ver o quão indefinido é o campo.

“Como acontece com todo patrimônio documental, os princípios de seleção podem variar de país para país, embora os principais critérios para se decidir quais os materiais digitais devem ser mantidos sejam sua importância e durabilidade cultural, científica, evidencial, ou qualquer outro valor. As decisões relativas à seleção e a quaisquer revisões subseqüentes precisam ser tomadas de maneira responsável, com base em princípios, políticas, procedimentos e padrões definidos” (UNESCO, 2003).

A partir da discussão de algumas realizações/iniciativas no cenário internacional poderemos identificar certas prioridades, certas estratégias de valoração do patrimônio digital ou, em alguns casos, a opção por uma preservação tão vasta que parece desprovida de critérios de seleção. Mas se neste aspecto da atribuição de valores, o patrimônio digital compartilha de uma problemática comum com o patrimônio em geral, do ponto de vista tecnológico, há especificidades que merecem atenção. A tecnologia digital impõe, por exemplo, novos desafios com relação à perenidade/permanência dos objetos, uma vez que os atuais discos rígidos trazem o prognóstico de no máximo um século de durabilidade. Esta e outras particularidades da tecnologia digital serão também discutidas ao abordarmos alguns casos de preservação do patrimônio digital.

Entre os diversos projetos voltados para o patrimônio digital, observamos duas vias distintas de atuação. Embora todas as realizações visem a preservação, os propósitos e, por conseguinte, as estratégias, configuram-se muito diferentes. Por um lado, temos os projetos que se propõem a passar para o formato digital acervos já existentes em outros meios. Por outro, há as iniciativas dirigidas aos arquivos que já nascem digitais, entre estes, os arquivos da Web.

Optamos por abordar cada uma destas vertentes, mapeando os seus principais desenvolvimentos para, em seguida, lançando mão de uma reflexão teórica, discutir tanto as possíveis questões em comum aos dois campos como também as suas problemáticas específicas.

1.1.1 Iniciativas relevantes de digitalização: Europeana e Biblioteca Digital Mundial¹

Entre as grandes iniciativas de digitalização de acervos, e sua posterior disponibilização, destaca-se, certamente, o portal Europeana, projeto empreendido pela Comissão Europeia, órgão executivo da União Europeia. Lançado em 2008, o portal se define como a biblioteca, o museu e o arquivo digital da Europa. A ideia de criar uma biblioteca digital de toda a Europa surgiu em 2005 e começou a ser implementada em 2007 num esforço conjunto de diversas instituições de memória europeias. Em 2014, a Europeana agrega os acervos digitalizados de 2.500 instituições, entre bibliotecas, museus, arquivos e galerias de 36 diferentes países. O gerenciamento do portal está situado na Koninklijke Bibliotheek, a Biblioteca Nacional da Holanda, em Haya.

A Europeana disponibiliza atualmente mais de 30 milhões de objetos digitais, relativos a textos, imagens, filmes, áudios e objetos tridimensionais². A maioria do conteúdo disponibilizado no portal é anterior ao século XX e já está em domínio público, mas há um parcela que permite o acesso apenas aos metadados. Quando uma instituição se torna parceira do projeto, não precisa necessariamente passar os direitos autorais do seu acervo. Cada instituição deve compartilhar os metadados, podendo escolher níveis diversos de acessibilidade ao conteúdo.

1 Uma parte das considerações tecidas neste item foram apresentadas em DANTAS; DODEBEI,2010.
 2 Dados disponíveis em: <http://pro.europeana.eu/documents/858566/640ac847-0dfc-4b01-9f36-d98ca1212ec9>. Acesso: 12/01/2014.

Imagem 1 - Página inicial do site Europeana em 2013

Home My Europeana Choose a language ▾

europeana think culture Search ▾ Search Help

Europa in een bibliotheek

Title:
Titelpagina voor: L'Europe savante, 1718

Description:
De vrouwelijke personificatie van Europa zit op een paard in een bibliotheek. In haar handen een scepter en een hoorn des overvloeds. Ze wordt vergezeld door putti met attributen die de kunsten en wetenschappen symboliseren. Op de achtergrond twee mannen in een werkplaats. Titelpagina voor: L'Europe savante. Den Haag: Alexandre de Rogissart, 1718.

Creator:
[Picart, Bernard](#) (Picart, Bernard) YYYY From: 1673-06-11— To: 1733-05-08; [Rogissart, Alexandre de](#) (Rogissart, Alexandre de) YYYY From: — To: -

Contributor:
[Bernard Picart](#) (Bernard Picart)

Coverage:
[first quarter 18th century](#); [eerste kwart 18e eeuw](#)

Geographic coverage:
[Amsterdam](#) ;; [The Hague](#) ;;

Public Domain marked

View item at [Rijksmuseum](#)

Fonte: Europeana (2013a)

O principal propósito da Europeana é fazer chegar a um público mais amplo a memória cultural europeia, encarnando no ambiente digital a noção de que o legado cultural pode contribuir para a construção de uma identidade europeia. A premissa é a de que “a culturally connected Europe is a better Europe” (EUROPEANA, 2014:19). Tal perspectiva está totalmente em sintonia com a tradição de transmissão de cultura encarnada pelas bibliotecas, museus e arquivos sendo, neste sentido, transposta para o suporte digital numa clara defesa da convergência entre as funções exercidas pelas instituições de memória (ROBINSON, 2014:414).

O discurso da Europeana em favor da acessibilidade do patrimônio como forma de promover o intercâmbio de conhecimentos e enriquecer a produção cultural está também vinculado a um propósito de incrementar a economia. A argumentação é que a difusão da herança cultural pode trazer benefícios econômicos relevantes, como está claro no seguinte trecho do plano estratégico da Europeana:

“We have carried out qualitative research and sought insight from established consultancy firms to calculate our expected economic impact and the investment required to make it happen. We have analysed our log files and cross-referenced them with user behaviour.

This has given valuable insight into where we can expect positive effects, or ‘impact’ from our actions. We now know that we are most effective when we take the lead in organizing solutions that concern a large part of the sector – such as copyright. And also, when we use our reach to develop big pan-European themes such as Europeana 1914-1918.(...) The largest economic contribution expected from Europeana is likely to be in uplifting tourism numbers and research quality.” (EUROPEANA , 2014:19)

O relatório *Europeana Strategy 2015-2020*, apresenta ainda um panorama do nível de digitalização das instituições de memória europeias, afirmando que 90% do que se encontra nos acervos europeus ainda não foi digitalizado. Esta defasagem requer, segundo o documento, um esforço contínuo para aumentar o montante de materiais digitalizados e, portanto, torná-los mais acessíveis. Mas é preciso, mais que tudo, fazer com que os 10% já digitalizados se tornem mais utilizáveis, o que inclui, por exemplo, uma preocupação com as formas de licenciamento e com a padronização dos metadados (EUROPEANA, 2014:9).

Outra iniciativa importante voltada para o objetivo de preservar e tornar acessíveis acervos digitalizados é a Biblioteca Mundial Digital, lançada em abril de 2009. Trata-se de um empreendimento de caráter internacional, com apoio da Unesco e financiamento das principais empresas de tecnologia do mundo, tais como Microsoft e Google. De caráter bastante seletivo, a Biblioteca Mundial Digital (BMD) propõe-se a disponibilizar acervos culturais já previamente qualificados como bens fundamentais em determinado país. Assim, são as instituições nacionais que selecionam os materiais que integram o acervo. O Brasil é representado pela Coleção D.Teresa Cristina Maria, com 142 itens cedidos pela Biblioteca Nacional³.

O projeto, de natureza educativa, patrimonial e filantrópica, encampado pela UNESCO em 2005, tem por objetivo, segundo depoimento de James H. Billington, diretor da Biblioteca do Congresso Americano e principal promotor da BMD, reduzir a lacuna digital entre os hemisférios norte e sul, ao tornar acessível um maior número de “documentos fundamentais” que pertencem à cultura planetária, tais como manuscritos, mapas e filmes. A principal característica da BMD é a seletividade de seu acervo, ou como afirma Billington, a hiper-seletividade, que a diferencia de outras bibliotecas online, como por exemplo a Europeana e a biblioteca do Google⁴. Há muito não encontrávamos na internet projetos que assumissem e defendessem a seleção no momento da ofer-

-
- 3 A coleção D. Teresa Cristina Maria foi doada pelo imperador D. Pedro II em 1891, após ser deposto. A coleção, que traz o nome imperatriz, inclui mapas manuscritos, curiosidades científicas e um vasto e significativo conjunto de fotografias, entre outros itens (BIBLIOTECA NACIONAL, 1987).
 - 4 Lançada em 2005, a biblioteca digital da Google Inc. disponibiliza 7 milhões de livros (GOOGLE BOOKS, 2010).

ta de objetos. O modelo de preservação patrimonial que vem sendo mais utilizado até o momento pode ser descrito como o do “excesso”, quer dizer, a prática de duplicar a totalidade de acervos e lançá-los na rede para favorecer a sua permanência por um período de tempo maior.

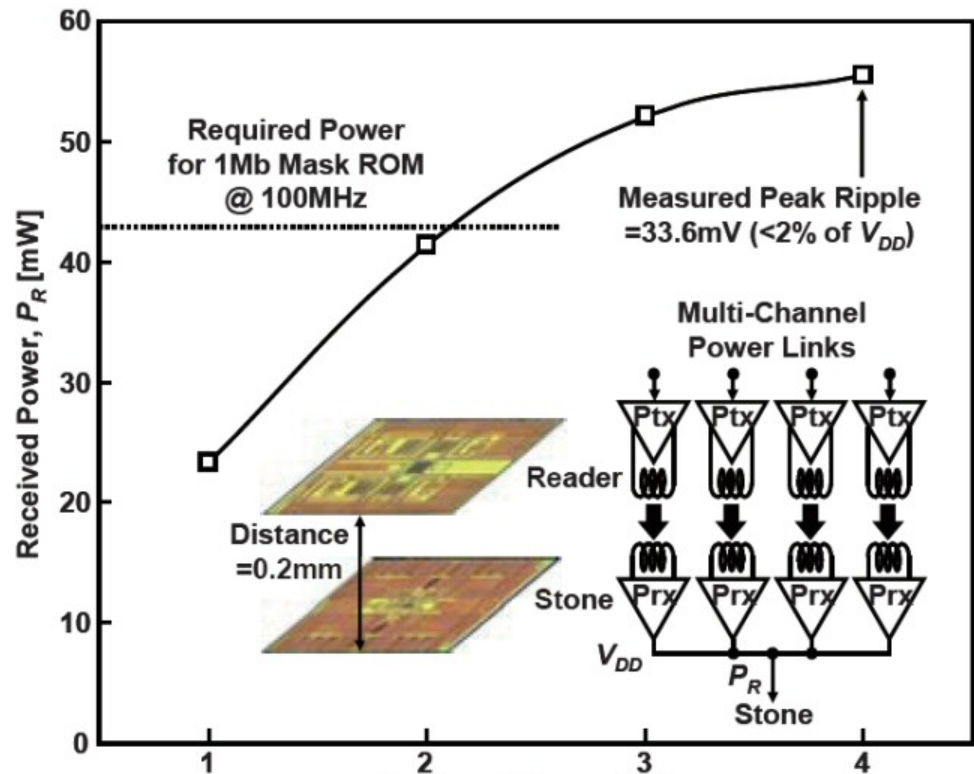
Evidentemente, a seleção, esta palavra-chave já tão conhecida dos processos de organização do conhecimento, implica custos muito elevados. Não foi diferente o que ocorreu com a criação da Biblioteca Mundial Digital. O projeto obteve apoio de instituições com um aporte de recursos de aproximadamente 10 milhões de dólares americanos. Importa notar que parte do financiamento do projeto destinou-se à inclusão de determinados materiais, como certos manuscritos científicos dos países árabes, que foram digitalizados graças à doação de um milhão de dólares feita pela Universidade de Ciência e Tecnologia King Abdullah, da Arábia Saudita.

O esforço de digitalizar o passivo da memória do mundo para disponibilizá-lo na rede mundial de computadores vem sendo questionado, principalmente pelo encargo financeiro dos direitos autorais. Bibliotecas digitais como a Europeana e a Google Books, que disponibilizam acervos protegidos pela legislação autoral, enfrentam ainda disputas judiciais para manter seus projetos em atividade. Com relação ao Google e sua estratégia de digitalização dos livros, há inclusive outras polêmicas em jogo. Robert Darton (2009) propõe uma crítica ao modelo em curso e defende que a gigantesca biblioteca digital do Google possa ser gerida não apenas pelas leis do mercado, mas também pelos interesses da sociedade civil. A seletividade com a qual a BMD se moldou, minimiza este problema da propriedade intelectual singular, pois transforma os objetos, em sua maioria já de domínio público, em coletivos da humanidade.

Outro aspecto importante a destacar na proposta da BMD é a preocupação com a questão da durabilidade. A biblioteca tornou-se a primeira cliente de uma nova tecnologia com capacidade estimada de preservar informações digitais por mil anos, a chamada *Digital Rosetta Stone*. As materialidades do registro digital ficam explicitadas neste artefato. Como afirma Wolfgang Ernets: “the ongoing existence of material memory agencies, both hardware and institutions, which still govern the power of deciding what can be stored legally and technically” (2013:97). Na ilustração abaixo vemos várias camadas de discos rígidos, selados com silício, e indicações de uma complexa tecnologia para impedir o contato direto com o mundo externo. De acordo com os pesquisadores da Universidade de Kioto, no Japão, as informações arquivadas neste artefato podem durar um milênio.

Imagem 2 - Herança cultural para futuras gerações

[VLSI] 'Digital Rosetta Stone' Passes Down Cultural Heritage to Future Generations



Fonte: Nikkei Technology:2009.

A proposta da Biblioteca Mundial Digital está em sintonia com o preceito de preservação patrimonial da modernidade. Uma vez inserida a informação nos discos de silício pode-se alcançar supostamente a segurança, mas perde-se a possibilidade de modificá-la. Desafios muito diversos se apresentam quando se trata de preservar os conteúdos da Web. No tópico seguinte expomos algumas importantes experiências em curso neste sentido, discutindo também os impasses e questionamentos que este novo objeto de preservação traz ao paradigma moderno do patrimônio.

1.1.2 Experiências de arquivamento da Web

Como preservar o material nascido digital, com seu caráter hipertextual e suas múltiplas formas de interação? Seria possível conservar algo do atributo fluido da interação na rede? Arquivar correspondências implica em determinados procedimentos arquivísticos já bastante corriqueiros. Porém arquivar mensagens eletrônicas ou *tweets* é um desafio que ainda está sendo enfrentado. Aliás, a questão da “gestão de documentos eletrônicos” é um eixo central nas discussões contemporâneas da Ciência da Informação e da Arquivologia (RONDENELLI, 2005).

Neste cenário, uma série de programas, livres ou comerciais, foram criados com vistas à gestão de documentos eletrônicos. O *Omeka* (www.omeka.org), por exemplo, foi criado a partir de um convênio com universidades americanas e trata-se de um software livre para gestão de documentos digitais. Este programa vem sendo utilizado por pesquisadores e por outras instituições que se propõem a criar acervos digitais com os registros do passado na rede. Há também, algumas invenções que se prestam a análises de documentos eletrônicos como o *PostHistory*, desenvolvido no MIT, um software que se propõe a organizar caixas de mensagens eletrônicas. O projeto desenvolveu-se a partir da constatação de que os nossos cotidianos *e-mails* constituem-se em fonte relevante para a pesquisa histórica e que ainda não dispomos de ferramentas adequadas para manejar um volume muito grande de informação em suporte eletrônico. É um projeto ainda em fase experimental que gera visualizações a partir de uma massa gigantesca de informação e pode trazer algumas inspirações para o pesquisador e para o usuário (VIEGAS et al., 2004).

É cada vez mais comum a criação de ferramentas que buscam “a possibilidade de indexar e recuperar os recursos informacionais com auxílio de programas lógicos que substituem o trabalho humano de representação dos objetos informacionais mas que permitem a leitura dos objetos tanto por humanos como por softwares” (DODEBEI, 2006:1). Há uma variedade de interfaces que representam propostas diversas de preservação e acesso aos dados de acordo com os objetivos institucionais e mesmo a legislação de direito autoral vigente em diferentes países. Abordaremos a seguir dois importantes projetos de preservação da Web – a Biblioteca Real da Holanda e o Arquivo da Web Portuguesa – iniciando com um breve histórico sobre o desenvolvimento do campo.

O arquivamento de páginas da web iniciou-se como uma ferramenta das empresas de busca ou de propaganda, que tinham como intenção construir bases de dados sobre a rede. O maior arquivo de páginas da internet foi criado em 1996, a partir da base de dados da empresa norte-americana Alexa. Brewster Kahle, sócio da empresa, fundou o Internet Archive com o claro in-

tuito de preservação das páginas. Tal experiência, que se tornaria uma matriz de formas de armazenamento e preservação dos conteúdos da web, será amplamente discutida no segundo capítulo.

O final da década de 1990 foi também o momento em que algumas bibliotecas e arquivos deram início às primeiras coleções de páginas da web. O final da década de 1990 coincidiu com a adoção por algumas bibliotecas e arquivos das primeiras coleções de páginas da web. A Biblioteca Nacional da Austrália foi uma pioneira, tendo fundado seu arquivo digital PANDORA (Preserving and Accessing Networked Documentary Resources of Australia) já em 1996. Com escopo seletivo, a biblioteca tem por objetivo a coleta de páginas da web australiana. A partir de 2001, a coleção já contava com uma ferramenta de busca por palavras – a PANDAS – permitindo assim que a maior parte do seu conteúdo seja consultada online. Mais da metade dos websites arquivados no PANDORA integram a rede de páginas do governo, o que aponta para os critérios envolvidos na seleção de websites, demonstrando a preocupação da instituição em preservar a produção digital relacionada ao Estado. A experiência da Biblioteca Nacional da Austrália é sem dúvida uma das pioneiras e hoje o acervo é fruto de uma colaboração com mais nove instituições do país. Em 2004, a coleção foi internacionalmente reconhecida, sendo inserida no registro Memória do Mundo da UNESCO.

Diversas instituições de salvaguarda cultural lançaram projetos específicos de arquivamento da web, mas que constituem modelos diversos e atendem a especificidades de cada país. A tabela abaixo aponta o avanço de uma multiplicidade de iniciativas que se expandem ao longo da primeira década do século XXI.

Tabela 1 - Instituições com projetos específicos de arquivamentos da web

Ano de criação	Nome	País
1995	University of North Texas CyberCemetery	USA
	Australia's Web Archive	Australia
1996	Internet Archive (provides Archive-it service)	USA
	Our digital island, a Tasmanian Web Archive	Australia
	Sweden (Kulturarw3)	Sweden

Ano de criação	Nome	País
1999	New Zealand Web Archive	New Zealand
2000	Bentley Historical Library (University of Michigan) Web Archives	USA
	Library of Congress Web Archives	USA
	WebArchiv (National Library of the Czech Republic)	Czech Republic
2001	Nasjonalbiblioteket	Norway
	National Library of Korea – OASIS (Online Archiving & Searching Internet Sources)	Korea
2002	Bibliotheca Alexandrina's Internet Archive	Egypt
	Japan Web Archiving Project	Japan
2003	Bibliotheksservice-Zentrum Baden-Württemberg	Germany
	UK Government Web Archive	UK
	Web Information Collection and Preservation - WICP (Chinese Web Archive)	China
2004	Croatian Web Archive (Hrvatski arhiv weba - HAW)	Croatia
	Internet memoryFoundation	France, Netherlands
	The UK Web Archive	UK

Ano de criação	Nome	País
	Government of Canada Web Archive (GCWA)	Canada
	National Library of Latvia	Latvia
	Netarkivet.dk	Denmark
2005	North Carolina State Government Web Site Archives	USA
	PADICAT: The Web Archive of Catalonia	Spain
	Web archive of the German Bundestag	Germany
	Web Archiving Service from California Digital Library (WAS service)	USA
	BnF - BnF Web Legal Deposit	France
	Finnish Web Archive	Finland
2006	Hanzo Archives	UK / USA
	Harvard University Library: the Web Archive Collection Service (WAX)	USA
	Koninklijke Bibliotheek	Netherlands

Ano de criação	Nome	País
2007	Central State Electronic Archives of Ukraine	Ukraine
	DILIMAG (Digital Literature Magazines)	Austria
	NTU Web Archiving System, NTUWAS	Taiwan
	Portuguese Web Archive	Portugal
	Slovenian Web Archive	Slovenia
	Stanford University Libraries	USA
	Tamiment Library and Robert F. Wagner Labor Archives at New York University	USA
	Web Archive Taiwan	Taiwan
2008	World Bank Web Archives	USA
	Basque Digital Heritage Archive	Spain
	Finnish Web Archive	Finland
	Web Archive Switzerland	Switzerland
2009	Web@rchive Austria	Austria
	Archive Team	Worldwide
	Archivo de la Web Española	Spain
	Columbia University Libraries Web Resources Collection Program	USA
	Ina (Institut National de l'Audiovisuel)	France
	University of Texas at San Antonio Web Archives	USA
	Web archive of Cacak	Serbia

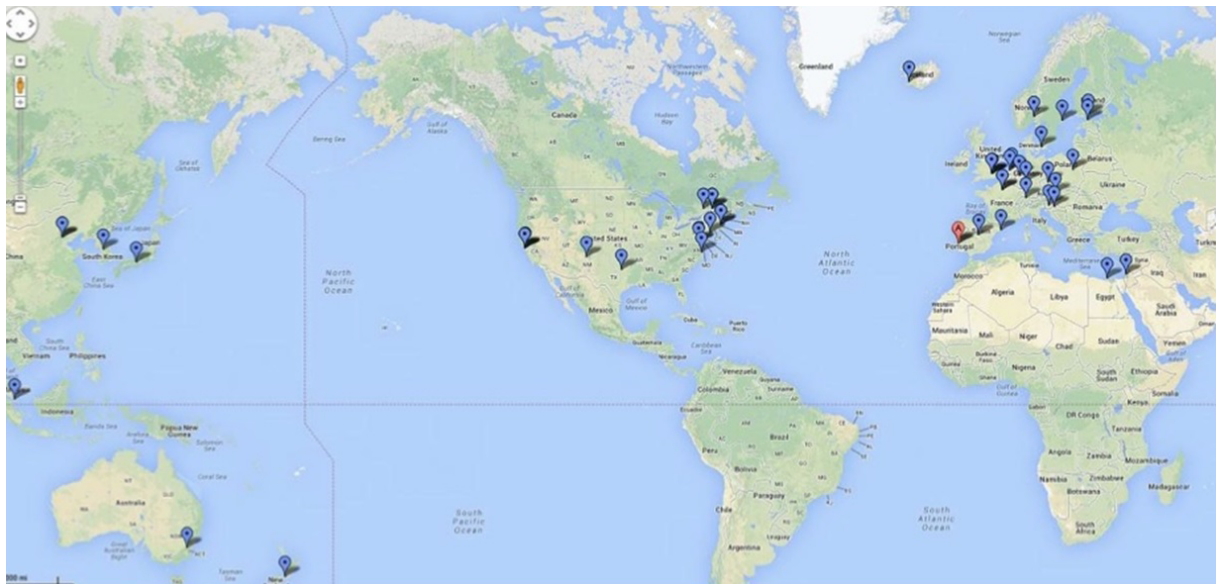
Ano de criação	Nome	País
	Aleph Archives	Switzerland / USA
	AUEB Web Archive	Greece
	E-diaspora (Télécom ParisTech, FMSH)	France
2010	Estonian Web Archive	Estonia
	OpenGovData Russia Archives	Russia
	qumram	Switzerland
	Reed Tech Archives	USA
	Cornell University Library	USA
2011	Internet Memory Research (ATN service)	France
	Israel Web Archive	Israel
	Archive.is	Worldwide
2012	Preservica	Worldwide
	Web Archiving Bucket	Switzerland / USA / Canada

Adaptado a partir das referências de GOMES et al (2011)

Dentre as iniciativas de preservação da Web, observamos basicamente dois tipos de proposta. Há os arquivos dos domínios nacionais, nos quais a principal diretriz é arquivar todas as páginas registradas em um determinado domínio, em geral de um país. É o caso, por exemplo, do Arquivo Nacional Português, que preserva todas as páginas do domínio .pt. E há os arquivos temáticos, cujo escopo é baseado na escolha de um tema. Cada tipo de arquivo implica em certas vantagens e desvantagens. O arquivo de largo escopo do domínio de um país, por exemplo, pode deixar de fora plataformas internacionais amplamente utilizadas naquele local. Por outro lado, uma coleção temática pode não incluir o contexto mais amplo das páginas selecionadas. A diversidade do escopo está, portanto, associada aos objetivos da coleção.

Além da multiplicação de coleções de páginas da web, houve, desde o começo do século XXI, a criação de associações internacionais com o objetivo de organizar e fomentar as atividades deste setor. Em 2003, foi criada a International Internet Preservation Consortium (IIPC), que além de disponibilizar documentação técnica, promove encontros anuais e grupos de trabalho sobre os desafios do arquivamento da web. Neste sentido, configura-se como um importante instrumento para a defesa de uma agenda comum, que inclui a interoperabilidade dos acervos e, portanto, o melhor acesso ao material arquivado.

Imagem 3 - Distribuição dos 43 membros do IIPC



Distribuição dos 43 membros do IIPC. Hartman, Cathy N. (2013)
Disponível em <http://netpreserve.org/about-us/members>

Dentre os propósitos do IIPC inclui-se o estabelecimento de protocolos e a divulgação do conhecimento relacionado ao arquivamento da web. A missão da organização é prover um fórum de troca de conhecimento buscando, ao mesmo tempo, tornar-se uma voz ativa na defesa da preservação dos conteúdos originários da rede mundial de computadores. No portal da organização são disponibilizados vários documentos importantes sobre o estado da arte da preservação de materiais nascidos digitais. Foi através de um grupo de trabalho técnico do IIPC que o formato WARC, inicialmente criado pelo Internet Archive, passou a ser recomendado internacionalmente e, finalmente, aceito com o selo de qualificação do ISO. Um dos principais objetivos do IIPC é assegurar o acesso ao material arquivado, como está claro no seguinte trecho:

“Like any other materials that libraries and archives collect, web archives are selected to complement existing collections and serve different goals. (...)Web archives are born-digital collections that require special software tools for their use. Researchers can view archived web sites page-by-page or whole collections can be processed as data, revealing broad characteristics of the collections. The organizations affiliated with the IIPC are committed to ensure that their web archive collections are preserved and made accessible for future researchers, historians and the public” (IIPC, 2013).

Há também, paralelamente às articulações internacionais, a criação de instituições em diversos países para gerir os processos de arquivamento digital. É o caso da Digital Preservation Coalition (DPC), constituída em 2001, com o objetivo de criar uma organização comum para as instituições do Reino Unido. A DPC se propõe a assegurar a troca eficiente de informações entre os seus membros propondo estratégias para o manejo do patrimônio digital. A proposta da instituição é garantir o acesso ao legado cultural registrado na web, exercendo assim um papel ativo na proposição de políticas públicas que favorecem a preservação do patrimônio digital. Além disso, a organização oferece treinamentos para profissionais que visam garantir as boas práticas no setor e organiza encontros onde são discutidas questões relacionadas ao quadro regulatório no qual se inserem as atividades de preservação digital. Uma outra forma de atuação do DPC é a distribuição de um Prêmio para as melhores iniciativas de preservação digital – o “Digital Preservation Awards”. De fato, a atuação do DPC extrapola os limites disciplinares e se propõe a prover discursos para uma audiência mais ampla, como fica explícito na passagem abaixo, que resume a proposta de uma de suas publicações:

“Mind the gap is a intended to win the hearts and minds of a much wider audience and to encourage all who are involved in creating and acquiring digital information, to seize the opportunity to develop a practical, coherent strategy for responsible management of the vast quantities of digital information being created and used in the UK today (DPC, 2006)”.

A criação de associações do gênero responde, de certa forma, a um clamor por ações de preservação digital que se iniciou na década de 1980, relacionado inicialmente à digitalização do patrimônio documental, se intensificando ao longo da década de 1990 (DAY, 1997). A questão da obsolescência das tecnologias digitais constituía o ponto central na caracterização de um cenário que exigia uma ação urgente. Com o passar do tempo, a preservação da Web passa a ser também incorporada como uma meta.

1.1.3 *A Biblioteca Real da Holanda e o Arquivo Nacional Português*

O arquivo digital da Biblioteca Real da Holanda (Koninklijke Bibliotheek), criado em 2008, é hoje uma das experiências de referência para o arquivamento da Web no cenário mundial. Mas é importante perceber que a sua implementação é resultado de uma série de lutas internas pela preservação dos arquivos nascidos digitais. De acordo com um relatório da Netherland Coalition for Digital Memory, órgão que agrega as instituições de memória holandesas, em 2009, a maior parte das instituições do país ainda se preocupava mais com a digitalização dos acervos do que com a preservação de registros nascidos digitais.

“...the cultural heritage sector’s main concern is still with physical collections. Digital records are mostly derivatives of physical collections: metadata or digital copies for online advertising. Where real (substitution) digitisation occurs, it takes up the available resources to such extent that planning beyond the phase of digitalisation is, for the moment, a bridge too far.” (NCDM, 2009:7)

Neste cenário, o relatório propõe que sejam desenvolvidas políticas públicas que possibilitem a preservação de longo termo do patrimônio nascido digital através da criação de novos acervos ou de desenvolvimento dos já existentes. Isso parece de fato ter sucedido com a Biblioteca Real, que se fortaleceu ao longo dos anos. Tendo iniciado com 100 páginas e contando atualmente com cerca de 5.100 páginas, o arquivo digital da Koninklijke Bibliotheek (KB) é uma coleção bastante seletiva, que considera no seu enquadramento aspectos relativos à língua, ao domínio (local de host) e à temática.

Um primeiro nível de seleção diz respeito diretamente à língua e ao domínio. Podem ser incluídos no arquivo websites que estão em holandês e registrados no domínio nl; websites em outra língua, mas registrados na Holanda; websites que estão em holandês, porém registrados em outro país; e websites estão em outras línguas, registrados em outros países, mas se referem à Holanda. Assim, por exemplo, um website de uma instituição internacional atuando na Holanda e com registro no domínio nl, poderia ser incluído no arquivo, como também uma página em inglês, hospedada nos EUA sobre a participação dos holandeses na Segunda Guerra mundial pode (WELTEVREDE, 2009:72). Mas há ainda critérios que consideram aspectos como o conteúdo dos websites, e estão baseados no sistema de classificação da Unesco.

The selection principles are based on the KB's general collection policy, which aims to store and make accessible items about the Dutch language, culture and society. (...). The primary selection stems from Websites with academic and cultural content, although innovative Websites as examples of current trends in the Dutch part of the Web are also considered (WELTEVREDE, 2009:74).

A seletividade da KB diz respeito não apenas à escolha dos websites, mas também às restrições legais que encontra. O arquivo só pode preservar páginas com a permissão legal. Além disso, a coleção digital só pode ser consultada na própria biblioteca devido às restrições relacionadas a direitos autorais no país. Sob a legislação holandesa, as páginas arquivadas são vistas como publicações impressas, colocando grandes desafios, como está registrado na apresentação do arquivo no portal da biblioteca:

“One important hindrance to reusing the data in the web archive must be mentioned here: the Copyright Act. The present Act (1912) is still based on printed information. Until 70 years after the death of an author/photographer etc. permission must be granted by the rights holders for any type of copying/reuse. And each version of a website may have many rights holders (photographer, journalist, designer, writers, etc.)” (KB, 2014).

A dificuldade de acesso à coleção é um problema que está sendo enfrentado com a elaboração de um acordo com a Universidade de Amsterdam, desde 2011, para a criação de ferramentas que possibilitem o acesso aos dados coletados. Em 2013, participei de um workshop no Digital Methods Initiative onde um grupo de pesquisadores e pós-graduandos convidados puderam testar e avaliar as novas formas de acesso propostas. As ferramentas que promovem um acesso aos metadados, mas não ao website arquivado, constituem-se numa estratégia para tornar o acesso possível em um contexto de restrição legal. Uma das propostas da biblioteca é criar um index que possa ser acessado via web, o que facilitaria o acesso ao material arquivado. Assim, o usuário poderia executar buscas por palavras no acervo, solicitar algumas estatísticas, encontrar um determinado endereço, porém não poderia ter acesso aquele website arquivado via web. Embora não esteja explícito nos termos oficiais do website, a equipe vem trabalhando para tornar acessível o conteúdo e uma das possibilidades é gerando dados que possam fazer o usuário encontrar aquele website em outros arquivos, principalmente o Internet Archive.

Imagem 4 - Página inicial do website *Wayback Machine*



Fonte: Internet Archive

Se o arquivo digital da BK é muito seletivo, seja pelos critérios institucionais ou pelas restrições legais, é importante também destacar que o caráter denso do seu acervo. Embora utilizando a tecnologia do Internet Archive, o modo de arquivamento da BK difere desta proposta. Diferentemente dos instantâneos ou snapshots coletados pelo Internet Archive, o arquivo da web nacional holandesa foi realizado a partir de um preceito de completude do material arquivado. Cada website selecionado como ponto de partida (seed) foi arquivado com todos os links. Assim, ao mesmo tempo que o escopo é muito seletivo, o fato da captura ter sido em vários níveis gera, paralelamente, um acervo não-intencional. Em um projeto piloto para investigar a dimensão deste arquivo não intencional, relativo apenas às páginas arquivadas em 2012, a equipe do arquivo digital, concluiu que:

“...the collection contains 12,327,673 unique pages, of which 11,395,072 were included on the seedlist (and 932,601 were only included unintentionally). The ‘aura’ that we uncovered contains an additional (unarchived!) 11,897,662 page representations. In other words, the uncovered web that is only indirectly collected while crawling consists of almost as many pages as the intentionally harvested collection!” (SAMAR, HUURDMAN et al., 2013:3).

Estas explorações do arquivo procuram também chamar atenção para a arquitetura da coleção digital. De acordo com Anat Ben-David, uma das coordenadoras do projeto, as ferramentas buscam tornar a própria busca um elemento importante da pesquisa (“search as research”) ampliando as possibilidades de conhecimento. O próprio arquivo busca favorecer a exploração dos dados e ampliar as possibilidades de pesquisa, procurando se tornar mais do que um repositório capaz de oferecer respostas a perguntas já estabelecidas pelos pesquisadores. A proposta seria a de ao propor outras formas de acesso, tornando possível um engajamento que passe a formular novas questões a partir do arquivo digital. Uma possibilidade proposta seria navegar na coleção através dos códigos postais citados nas páginas e, portanto, aproximar-se da coleção a partir de um critério espacial.

Pelo que foi exposto, podemos concluir que a proposta da Biblioteca Real da Holanda é a de construir uma coleção pequena, porém densa e com vista a uma preservação de longo prazo. Assim, embora utilize ferramenta criadas pelo Internet Archive, o propósito da coleção é bastante distinto. Enquanto o Internet Archive preserva os websites parcialmente e apenas em dois níveis, a KB se propõe a salvar todos os links mencionados e todas as imagens de um conjunto pequeno de websites escolhidos segundo os critérios institucionais. Esta experiência chama atenção para os modos distintos como um mesmo artefato técnico, neste caso o Heritrix desenvolvido pelo Internet Archive, e um mesmo tipo de formato, pode ser reprogramado para atender a demandas distintas e concepções de preservação dissonantes.

O Arquivo da Web Portuguesa (AWP) é uma outra iniciativa de preservação de web nacional muito relevante no cenário atual, pois é um dos poucos países que dá acesso online ao conteúdo arquivado. O Arquivo foi criado em 2008 pela Fundação de Computação Científica Nacional (FCCN) e hoje conta com 1200 milhões de objetos digitais arquivados (entre 1996 e 2012). O acervo anterior à data de criação do AWP foi transferido do Internet Archive com o objetivo de tornar o acesso possível. Um dos motes da equipe do arquivo é o provérbio português “guarda o que não presta, acharás o que precisas”, que explica a opção por uma arquitetura de preservação que inclui todas as páginas do domínio do país. Eis a definição do escopo da coleção digital do AWP:

“1. Está sob a hierarquia .PT; 2. Não está sob a hierarquia .PT mas o conteúdo está embebido numa página alojada sob .PT. O objectivo é recolher a informação necessária para que as páginas portuguesas arquivadas possam vir a ser apresentadas de forma completa. Por exemplo, uma imagem alojada no sítio web de partilha de fotografias www.flickr.com que esteja embutida numa página de um sítio web de blogs como, por exemplo, o blogs.sapo.pt, será considerada como parte da Web portuguesa; 3. Não está

sob a hierarquia .PT mas existe um redirecionamento a partir de um nome sob .PT. Por exemplo, uma firma multinacional que regista o domínio .PT com a sua marca mas que o coloca a apontar para o sítio web principal da empresa que está sob a hierarquia .COM.” (GOMES; MIRANDA, 2008:2)

Na citação acima percebe-se uma preocupação em executar operações de arquivamento que possam garantir uma certa contextualização das páginas arquivadas através da inclusão dos links referenciados nas páginas registradas no domínio português. Além da busca por palavras no acervo, o Arquivo da Web Portuguesa também se propõe a chamar atenção para alguns conteúdos, como as páginas que estão listadas na imagem abaixo sobre: José Saramago, EXPO 98, Luís Figo e UEFA Euro 2004. Estas referências não constituem uma coleção temática, mas sim o acesso a um único endereço arquivado pela instituição.

Imagem 5 - Área de pesquisa do site Arquivo da Web Portuguesa



ARQUIVO DA WEB PORTUGUESA

[Pesquisa avançada](#)

 **Pesquise e aceda a páginas do passado**
Veja ou reveja páginas que já desapareceram.

São mais de 1200 milhões de conteúdos arquivados entre 1996 e 2012 que estão ao seu dispor.

[Conheça o projeto](#)

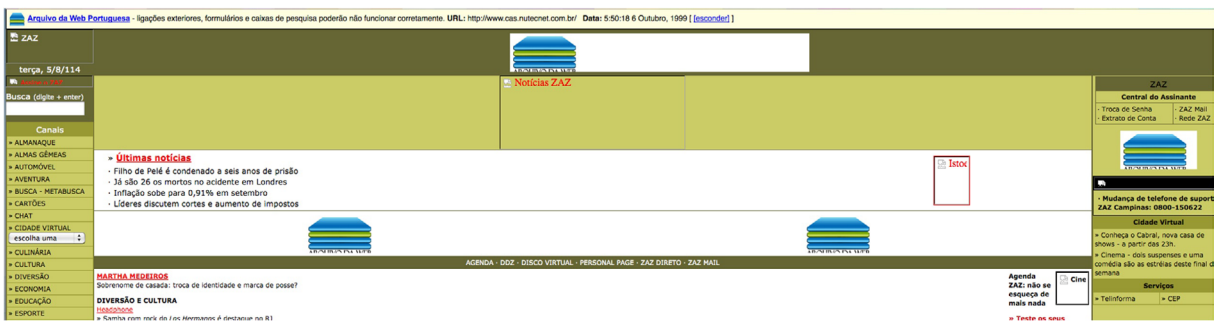
Alguns momentos relevantes registados no arquivo
Navegue por momentos que ficaram para a história.

 <p>José Saramago Prémio da Literatura 1998 Página pessoal em 2008</p>	 <p>Expo '98 Exposição Mundial de 1998 em Lisboa Página pós-evento</p>
 <p>UEFA Euro 2004 Campeonato Europeu de Futebol 2004 em Portugal Web blog</p>	 <p>Luís Figo Figo eleito melhor jogador do mundo Notícia do jornal "O Jogo" em 2001</p>

Fonte: http://sobre.arquivo.pt/?set_language=pt

Uma das vantagens do arquivo é a disponibilização dos conteúdos online, o que demonstra uma preocupação com o acesso às páginas. Por outro lado, a consulta ao acervo online traz uma série de especificidades relativas à atualização dos elementos dinâmicos de cada website. Esta questão será explorada ao tratarmos da Wayback Machine, software desenvolvido pelo Internet Archive, no terceiro capítulo. Abaixo a imagem mostra como uma página pode ser tomada por alguns elementos contemporâneos, neste caso o logo do Arquivo da Web Portuguesa. Na verdade, isto demonstra ser uma estratégia na atualização dos banners publicitários, porém não deixa de ser a retirada de um conteúdo que pode ser importante para alguns pesquisadores.

Imagem 6 - Não há registro de publicidade nos sites arquivados no Arquivo da Web Portuguesa



Fonte: Internet Archive

A limpeza destes banners pode por exemplo desapontar um pesquisador que esteja, por exemplo, interessado em estudar a história da publicidade na web. Mas, independente do tópico de pesquisa, a retirada dos banners exclui um elemento importante do modelo de navegação e minimiza a importância do mercado na configuração dos conteúdos online. A institucionalização dos dados, ou melhor das páginas arquivada, parece conter neste caso, de maneira sutil, uma tentativa de enquadrar essa memória documental tornando-a, em certo sentido menos mundana. Neste sentido cabe aqui a clássica assertiva de Terry Cook ao chamar atenção para o papel do arquivista na criação do corpus documental, afirmando a que ainda é necessário um esforço acadêmico para tornar visível os modos de criação dos registros⁵

⁵ Nas palavras de Terry Cook: “very little notice is still paid to how the record is chosen and shaped, privileged and marginalized, by archivist’s interventions” (2002:174)

1.2 Discursos de patrimonialização e novos conceitos no Brasil

As próximas páginas se debruçam sobre a emergência do conceito de memória digital no Brasil. Acompanhar de que modo as discussões estão se realizando e sob que principais referências é a intenção desta parte. Intriga-nos o pouco interesse em relação aos arquivos nascidos digitais, alvo de nossa pesquisa. Ao mesmo tempo que pretendemos pensar as criptografias da memórias, sobretudo nos arquivos internacionais já mencionados, pensamos ser relevante acompanhar as trilhas contemporâneas desta discussão no Brasil.

1.2.1 Noções de memória digital no Brasil a partir do estudo de caso do Fórum da Cultura Digital

Partimos da constatação de que se houve na experiência brasileira da internet uma rápida incorporação de redes sociais como Orkut e Facebook e de ferramentas de microblogging como o Twitter, no que se refere ao uso compartilhado de ferramentas de preservação da memória da web, com utilização de softwares de arquivamento, não é possível identificar experiências relevantes. Há muitos arquivos online que disponibilizam os seus materiais digitalizados, porém não há no país experiências de grande porte relacionadas ao patrimônio nascido digital (DANTAS; DODEBEI, 2010)⁶. O caso do Museu da Pessoa deve ser mencionado, uma vez que inclui em seu acervo a possibilidade do testemunho digital. Porém, trata-se, de fato, de um acervo misto composto de entrevistas e histórias de vida (áudio e vídeo), que só posteriormente passaram para o formato digital.

Desde o início da pesquisa intrigava-nos as poucas menções, nos documentos oficiais do país e mesmo nas discussões públicas que mapeamos, aos arquivos nascidos digitais e, mais especificamente, ao arquivamento da Web. Em uma reportagem de 2013⁷, intitulado “Quem vai arquivar a internet para futuras gerações?” encontra-se uma breve descrição deste cenário:

6 Parte da discussão realizada neste tópico encontra-se publicada no artigo Notes on the emerging concepts of born-digital heritage in Brazil. (DANTAS; DODEBEI,2010)

7 A reportagem foi publicada no Estadão do dia 24 de agosto de 2013 e encontrase disponível online em: <http://web.archive.org/web/20140917142057/http://acervo.estadao.com.br/noticias/acervo,quem-arquivara-a-internet-para-as-futuras-geracoes,9215,0.htm>

“Com 20 anos de internet, o Brasil ainda não conta com nenhuma instituição, legislação, diretriz ou esforço que garanta que conteúdos produzidos na web sejam vistos como patrimônio cultural e, assim, coletados, catalogados e armazenados. Símbolo da relevância digital deste começo de século 21, a memória das recentes manifestações de junho, paradoxalmente, só está garantida nos formatos impressos.” (ENTINI, 2013)

Diante deste cenário, interessou-nos investigar um pouco mais detidamente como as noções de memória e de patrimônio digital estavam se constituindo nesse momento no país; quais políticas públicas, programas ou debates envolvendo estas questões tinham lugar no cenário nacional. Enquanto nos perguntávamos como iríamos pesquisar essas questões, tomamos conhecimento da existência do Fórum de Cultura Digital, iniciativa do Ministério da Cultura (Minc) em cooperação com a Rede Nacional de Pesquisa (RNP)⁸, que então se iniciava. O Fórum nos pareceu uma rede importante de debates pelo fato de agregar tanto as vozes da esfera governamental como também de especialistas, ativistas e do público em geral, apresentando-se assim como uma via possível para mapear as posturas em voga com respeito ao patrimônio digital. Além disso, o Fórum nos interessou também pelo fato de se desenrolar privilegiadamente numa plataforma digital aberta, o que possibilitava uma participação efetiva nos debates, uma incursão etnográfica. A seguir, apresentamos o desenvolvimento deste percurso investigativo que, partindo do Fórum, buscou traçar também articulações com outras instâncias.

O Fórum da Cultura Digital Brasileira constitui certamente uma das primeiras iniciativas de abrir um campo de discussão sobre o patrimônio digital no país. Criado no ano de 2009, o Fórum tinha como proposta recolher uma gama de contribuições e fomentar uma rede de discussão em torno de questões cruciais para o país relacionadas à cultura digital. No livro *Cultura digital.br*, que traz uma série de entrevistas com pesquisadores e profissionais envolvidos com a cultura digital, visando incitar as trocas do Fórum, encontramos uma apresentação das diretrizes que nortearam a sua estruturação.

“O propósito do Fórum é articular os cidadãos brasileiros para a construção de políticas públicas democráticas de cultura digital. Eventos presenciais ocorrerão durante todo o segundo semestre de 2009. O diferencial, no entanto, consiste na utilização de um site de rede social para organizar os debates no plano virtual. Nes-

8 A Rede Nacional de Pesquisa foi criada em 1989, pelo Ministério da Ciência e Tecnologia, para constituir a primeira rede de conexão à internet no Brasil, voltada para a comunidade acadêmica. Atualmente agrega mais de 800 instituições de ensino e pesquisa no país, atuando em diversos projetos de educação, ciência e tecnologia.

se ambiente, os interagentes podem organizar sua rede política de relacionamentos, construir grupos e fóruns de discussão, propor atividades e conversas, produzir um blog, entre outras atividades” (SAVAZONI, 2009:8).

É importante situar o Fórum da Cultura Digital Brasileira no contexto de um investimento mais amplo do Ministério da Cultura (MinC), voltado para as mídias digitais. Como mostra Eliane Costa (2011), este investimento se inicia com a gestão de Gilberto Gil, entre janeiro de 2003 e julho de 2008, durante a qual as tecnologias digitais passam a assumir um papel crucial, consideradas como instrumento capaz de fomentar os campos da cultura e da educação e de ampliar o acesso a estes.

(...) o Ministério da Cultura identificou na revolução das tecnologias digitais um fenômeno de natureza essencialmente cultural, passando a sinalizar que o uso pleno desses recursos, calcado em políticas públicas, seria estratégico para o país, pela perspectiva de transformação relacionada à democratização do acesso ao conhecimento, ao estímulo de uma cultura de rede, à ampliação das possibilidades de produção de conteúdos culturais em mídia digital, e à sua posterior difusão na internet, sob a ótica dos direitos culturais e da diversidade. (...) Ao dar início à formulação de políticas públicas para o campo da cibercultura, o Ministério passou a identificá-las como políticas de Cultura Digital, tornando esse tema transversal em sua atuação (COSTA, 2011:116)

A atenção ao papel da cultura digital se fez presente em múltiplas ações da gestão de Gilberto Gil no MinC. A primeira política implementada pelo Ministério, em 2004, por exemplo – os Pontos de Cultura – que atuava apoiando iniciativas culturais já existentes, incluía a implantação de estúdios digitais com o objetivo de possibilitar a produção audiovisual e a sua difusão (COSTA, 2011:181). Todos os editais abertos pelo Ministério a partir de 2005, incluíam um aspecto relativo à cultura digital, alguns deles referindo-se especificamente a iniciativas de criação ou digitalização de acervos (FERNANDES, 2010).

Embora o Fórum de Cultura Digital e outros eventos que iremos abordar tenham se desdobrado a partir de 2009, ano em que o Ministério da Cultura já estava sob a gestão de Juca Ferreira⁹, é importante perceber que eles estão vinculados a todo um conjunto de iniciativas voltadas para a cultura digital que resultam de um processo já em curso. Isso é reconhecido pelo próprio Juca Ferreira que, por ocasião do primeiro seminário presencial realizado pelo Fórum,

⁹ Abordaremos também alguns eventos que têm seqüência nas gestões de Ana de Holanda (janeiro de 2011 a setembro de 2012) e de Marta Suplicy (setembro de 2012 até o momento) no Ministério da Cultura.

o define como “um momento especial, de consolidação das ações que começaram ainda na gestão de Gilberto Gil, quando o ministério se abriu para as discussões sobre novas tecnologias” (AGUSTINI, 2009). A preocupação do MinC com a cultura digital, que se estabelece na gestão de Gilberto Gil, se expressando em muitas de suas políticas, opera um corte em relação aos posicionamentos anteriormente adotados pelo Ministério. É, como observa Eliane Costa (2011), uma preocupação “não natural”, já que historicamente as tecnologias digitais vinham sendo tratadas – no âmbito governamental – prioritariamente em seu aspecto técnico e delegadas a instâncias políticas ligadas à Ciência, à Tecnologia e à Comunicação. Este posicionamento teve continuidade na gestão de Juca Ferreira que, ao comentar a importância do debate público em torno da cultura digital realizado pelo Fórum, afirma:

“Essa é uma questão pública vital, pelo potencial que a tecnologia digital, a internet e todo esse mundo novo que se abre com este desenvolvimento tecnológico, tem para transformações na sociedade. Isso vai ter implicações em todas as áreas da vida social e também nas características de um novo Estado que enfrente o século 21, que pense outras formas de se fazer política. É preciso atrair para o debate os segmentos direta ou indiretamente ligados a esta realidade. Esse é um papel pioneiro, mas importantíssimo. Uma função seminal de difundir os temas e retratar o nível em que estão sendo tratadas no mundo” (FERREIRA *apud* AGUSTINI, 2009)

É em meio a esse contexto mais amplo de um investimento do Minc na cultura digital que surge o Fórum da Cultura Digital Brasileira, estruturado em cinco eixos temáticos: Comunicação Digital, Memória Digital, Arte Digital, Infra-estrutura da Cultura Digital e Economia da Cultura Digital. O Fórum não se constituía como entidade legal capaz de promover ações efetivas, porém os documentos propostos a partir dele eram encaminhados às instâncias governamentais, podendo se tornar esboços de futuras políticas públicas. Como explica Rogério Lourenço (2010), curador do eixo Memória Digital, a ideia do Fórum era encurtar a distância entre a sociedade, com as suas demandas relacionadas à cultura digital, e as políticas desenvolvidas pelo Ministério da Cultura. O Fórum desempenhou um papel relevante, por exemplo, na discussão do Marco Civil da Internet, atuando como espaço de debates e também como ferramenta para a sistematização de contribuições enviadas ao projeto de lei¹⁰. Além das discussões pela plataforma da internet, o Fórum produziu também encontros presenciais como o Seminário Internacional do Fórum da Cultura Digital

¹⁰ Mais informações sobre esse processo podem ser acessadas no site do Fórum. Disponível em <http://culturadigital.br/>. Acesso: 05/04/2011.

Brasileira (novembro de 2009) e o II Fórum da Cultura Digital Brasileira (novembro de 2010), realizados na Cinemateca de São Paulo. Estes seminários incluíam mesas de debates com pesquisadores, profissionais, ativistas e artistas do Brasil e de outros países. E havia também a realização de plenárias de cada eixo temático, nas quais eram sistematizadas as discussões que vinham se desenrolando no Fórum da internet. Os seminários eram palco também de posicionamentos políticos, como foi o caso de um manifesto (“Carta em defesa da liberdade na internet”), articulado no segundo seminário, contra a o substitutivo PL89/83, que propunha uma série de controles sobre o uso da internet¹¹.

A maior parte dos membros do Fórum pertencia à região Sudeste, principalmente ao Estado de São Paulo, mas havia participantes das diversas regiões do país¹². Na figura abaixo é possível visualizar uma das páginas do Fórum, com alguns dos 239 grupos que o compunham em julho de 2011, abordando temas variados, tais como comunicação eletrônica, cibercultura, música na web, mídia livre etc. Até Julho de 2012, quando foi transformado numa “plataforma pública de blogs”, o Fórum chegou a ter cerca de 400 grupos, 500 fóruns e mais de 7000 participantes¹³.

11 Em nome do combate a práticas criminosas por meio da internet, o substitutivo PL89/83, encampado pelo Senador Eduardo Azeredo, continha artigos que implementavam uma série de controles sobre o uso da internet. A Carta do Fórum de Cultura Digital, assim como outros movimentos na época, considerava que os controles propostos, tais como “a vinculação de um número IP a identidades civis” implicavam numa “supressão de direitos e garantias individuais”, e poderiam cercear a criatividade e instituir um “vigilantismo que é democraticamente insustentável”. (Disponível em <http://culturadigital.br/documentos/>. Acesso: 22/04/2013).

12 Ver o mapa com a distribuição geográfica dos membros do Fórum em Janeiro de 2010. Disponível em: <http://culturadigital.br/blog/2010/01/19/culturadigital-br-um-mapeamento-dos-usuarios-da-rede/>. Acesso em 10/07/2010.

13 Dados retirados do site do Fórum. Disponível em <http://culturadigital.br/sobre/>. Acesso: 02/09/2013.

de que muitas discussões exigiam “uma inserção mínima no campo” e conhecimentos específicos como no caso dos fóruns em torno dos padrões de preservação e metadados.

Imagem 8 - Página de apresentação do grupo de Memória Digital



Página de apresentação do grupo de Memória Digital. Acervo da autora.

Os debates realizados no grupo de Memória Digital tiveram como foco principal a digitalização de acervos, incluindo desde aspectos técnicos até discussões mais amplas em torno, por exemplo, da necessidade de garantir e ampliar o acesso àquilo que é digitalizado. Esta preocupação com a digitalização dos acervos é também o ponto central do documento de base para o Eixo Memória Digital, lançado em novembro de 2009. Este texto é fruto de uma série de encontros do curador do Eixo, Rogério Lourenço, com os diversos órgãos do MinC e com a RNP, de discussões geradas em seminários como o que foi organizado pelo Comitê Gestor da Internet no Brasil (CGI.br)¹⁶, e dos debates já iniciados no Fórum de Cultura Digital. O documento apresenta um diagnóstico de todo um espectro de questões relacionadas ao patrimônio digital no país, incluindo os aspectos infra-estrutural, tecnológico, legislativo e cultural. O texto não se propõe a uma análise aprofundada das questões, mas se coloca como um mapeamento do estado atual do campo (LOURENÇO, 2010), elencando um conjunto de discussões e ações prementes para o seu desenvolvimento.

Citando o Plano Nacional de Cultura, aprovado em setembro de 2009 na Comissão de Educação e Cultura da Câmara dos Deputados, o documento defende a digitalização dos acervos como uma via importante para democratizar o patrimônio cultural.

¹⁶ O Seminário para formulação de estratégias de fomento à produção e disponibilização de conteúdos culturais digitais em língua portuguesa aconteceu em agosto de 2007, no Rio de Janeiro.

“Uma política de digitalização de acervos é indispensável para permitir uma circulação inédita de registros de nossa memória cultural, pictórica, gráfica e textual. (...) A convergência digital representa o ambiente contemporâneo de circulação da cultura, que deve ser observado sob uma perspectiva atenta à distribuição das tecnologias e às suas formas de utilização, bem como ao desenvolvimento de conteúdos, digitalização de acervos públicos e incentivos aos projetos experimentais”¹⁷ (PLANO NACIONAL DE CULTURA *apud* LOURENÇO, 2009)

Com relação à questão *o que deve ser alvo de preservação*, tanto o documento como as discussões no grupo de Memória Digital estavam alinhados com uma perspectiva antropológica de cultura, em consonância com a visão do MinC. Havia, portanto, uma preocupação com a preservação e disponibilização de uma multiplicidade de manifestações culturais brasileiras, para muito além da esfera erudita.

Além de apontar para a necessidade de elaboração de um Plano Nacional de Digitalização, com linhas de financiamento, o documento de base para o Eixo Memória Digital traça um panorama das experiências de digitalização já realizadas ou em curso no país. Havia na época, segundo o levantamento, tanto experiências relevantes – caso, por exemplo, do trabalho desenvolvido pela Biblioteca Nacional – como processos de digitalização realizados com tecnologias e metodologias inapropriadas. Mas mesmo os projetos bem sucedidos sofriam de uma falta de articulação:

“(...) os exemplos bem sucedidos de projetos de digitalização carecem de uma agenda e um diálogo sistematizado, tanto por especificidades, quanto por interesses e necessidades comuns. Não há, afora encontros acadêmicos ou informais, fóruns de discussão estáveis que reúnam as diversas iniciativas tecnológicas (científicas) e culturais, no que diz respeito à convergência de conteúdos e a troca de experiências de desenvolvimento de hardware e software. O Fórum de Cultura Digital é uma tentativa” (LOURENÇO, 2009).

Uma das medidas fundamentais para promover uma articulação dos múltiplos esforços de digitalização e o acesso ao patrimônio digitalizado seria, segundo o documento, a padronização dos processos de digitalização no país. Em 2010, o Conselho Nacional de Arquivos (Conarq) já havia lançado uma resolução (nº31) estabelecendo normas para a digitalização de documentos, mas esta não trazia especificações relativas, por exemplo, à digitalização de imagens em movimento. Essa necessidade da criação de protocolos para preservação da “memória digital” foi

¹⁷ O texto final do Plano Nacional de Cultura, instituído em dezembro de 2010, através da Lei 12.343, mantém a afirmação da necessidade de uma política nacional de digitalização.

também um tema bastante discutido no Fórum. Reproduzimos abaixo a postagem de um dos participantes do grupo de Memória Digital, Leo Germani, que diz respeito a esta questão.

“Tenho acompanhado alguns esforços aqui em São Paulo em se organizar acervos culturais em ambientes digitais. (...)”

Imagine que beleza, fazer uma busca por “Mário de Andrade” e ter como resultado todos os itens, sejam áudios, fotos ou textos, relativos a ele, com a indicação do local onde o item se encontra, e com uma versão para visualização imediata.

Imagine isso a nível nacional, e não só buscando itens do acervo de uma instituição específica, mas agregando acervos dos mais diversos espaços culturais.

O desafio é grande e tem muita gente quebrando a cabeça para tentar resolvê-lo, inclusive eu.

Definir um padrão único de banco de dados a ser seguido, uma unidade nos metadados é uma questão bastante complicada. Muitos acervos usam seus próprios métodos, outros não tem estrutura para implementar sistemas tão complexos. (...)”¹⁸.

18 Mensagem postada no grupo Memória Digital do Fórum Cultural Digital no dia 20 de agosto de 2009. Acervo da autora.

Imagem 9 - Fragmento da postagem de Leo Germani no grupo de Memória Digital

CULTURA DIGITAL
PLATAFORMA PÚBLICA DE BLOGS E CONVERSAS

Página Inicial Sobre Conceito Acervo

Por um padrão para os acervos digitais de cultura – uma proposta

Home > Forums > Cultura Digital > Curador de Memória Digital – Fórum > Por um padrão para os acervos digitais de cultura – uma proposta

Tagged: acervo, busca, digitalização

This topic contains 10 replies, has 7 voices, and was last updated by Fernão Lopes 4 years, 9 months ago.

Viewing 11 posts - 1 through 11 (of 11 total)

Author	Posts
20/08/2009 at 2:49	#4366
 Leo Germani <i>Member</i>	<p>Tenho acompanhado alguns esforços aqui em São Paulo em se organizar acervos culturais em ambientes digitais. A Secretaria de Cultura da cidade está fazendo um grande levantamento para escolher uma forma adequada de arquivamento para adotar em todos os seus equipamentos culturais.</p> <p>Imagine que beleza, fazer uma busca por "Mário de Andrade" e ter como resultado todos os itens, sejam áudios, fotos ou textos, relativos a ele, com a indicação do local onde o item se encontra, e com uma versão para visualização imediata.</p> <p>Imagine isso a nível nacional, e não só buscando itens do acervo de uma instituição específica, mas agregando acervos dos mais diversos espaços culturais.</p>

Fragmento da postagem de Leo Germani no grupo de Memória Digital no dia 20 de agosto de 2009.
Acervo da autora.

O participante dá seqüência à sua fala, propondo como uma possível solução para coordenar os dados de diferentes acervos culturais a utilização de um padrão XML – o RSS (Really Simple Syndication) – muito usado em sites de notícias. Este padrão permite, por exemplo, conectar blogs e sites de notícias que usam tecnologias bastante diferentes¹⁹. Esta e outras diferentes propostas surgidas no grupo de Memória Digital convergiam em torno da necessidade de criar protocolos para os acervos digitais. De fato, sem esta condição, torna-se impensável uma comunicação efetiva entre as diversas instituições do país, correndo-se o risco de desperdiçar recursos públicos, seja digitalizando duplamente um mesmo manuscrito ou reprocessando dados de um mesmo acervo em código binário. Embora uma centralização das políticas públicas relacionadas à preservação do patrimônio digital não seja desejável, essa reflexão sobre os


¹⁹ A mensagem completa enviada por Leo Germani, com a descrição da sua proposta, assim como o comentário de outros membros do grupo, integra o acervo pessoal dos registros desta pesquisa.

meios possíveis de regulação dos processos e uso eficiente dos recursos nos parece plenamente pertinente (DANTAS; DODEBEI, 2010).


É importante atentar também para o fato de que a criação de padrões para os acervos digitais não constitui uma questão meramente técnica, mas envolve uma série de interesses econômicos e políticos, aspectos também levantados no grupo. A discussão em torno da escolha de softwares livres ou de tecnologias proprietárias para os acervos digitais é parte dessa dimensão político-econômica das opções tecnológicas. Embora desde 2003 o uso do software livre fosse uma diretriz do Governo Lula, em 2009 esta meta ainda não constituía uma realidade. Embora algumas das instituições empreendessem um esforço nesse sentido, o software livre não era adotado de modo sistemático na totalidade das iniciativas de digitalização de acervos públicos [falta colocar nota falando de como se desenrolou até hoje essa questão do uso de software livre]. Tanto nas discussões do grupo como no documento de base do Eixo Memória Digital, o problema do estabelecimento de protocolos de arquivamento digital esteve estreitamente ligado a um posicionamento político em defesa do uso de formatos livres como forma de garantir o acesso à memória cultural digitalizada. É importante ainda destacar que toda essa discussão se realizava em referência aos acervos digitalizados, como vimos no depoimento citado, em que o exemplo dado são os documentos relacionados a Mario de Andrade.


A partir dessa incursão nas discussões do Fórum e no documento do Eixo Memória Digital, encontramos elementos para pensar o que se entendia por “memória digital” e de que modo esta noção estava sendo construída no âmbito institucional brasileiro entre o final da primeira e o início da segunda década do século XXI. A centralidade que a questão da digitalização de acervos ocupou tanto no Fórum como no documento do Eixo, se não explica o porquê, ao menos nos indica como a noção de memória digital constituída nesse período pouco levava em conta os arquivos nascidos digitais e negligenciava completamente os arquivos da web. Ao levantarmos no Fórum a questão dos arquivos nascidos digitais, verificamos que o tema não estava em pauta, embora tenha sido bem acolhido como uma área a ser futuramente explorada. Abaixo, a nossa primeira intervenção nesse sentido, junto com a resposta recebida.


Imagem 10 - Fragmento de página do Fórum com a nossa primeira intervenção


 [Carlos Seabra](#) disse 8 meses, 3 semanas atrás:
Mais algumas **questões adicionais** e complementares à discussão de memória digital e políticas relativas a acervos digitais:

- 1) Como fomentar que o usuário seja também produtor e alimente acervos de conteúdos alternativos, estimulando e mobilizando protagonismo social e popular (oportunidade de novos conteúdos e desafio de articulação).
- 2) Articulação de instituições e agentes: Não basta a disponibilização dos conteúdos: é necessário pensar nos seus usos e mediar/facilitar /estimular esse processo.
- 3) Contrapartidas sociais na forma de, conteúdos digitais livres, para projetos com financiamento total ou parcial com verbas públicas.
- 4) Direito de uso justo: questão a ser colocada desde já em discussão pública mobilizadora e articuladora, aproveitando as mudanças que estão sendo propostas em direitos autorais.
- 5) Fomento à visão de patrimônio social e cultural de peso global do nosso idioma, em ação partilhada com demais países e comunidades de lusofalantes. Ação complicada e que até hoje quase nada rendeu, mas não menos importante por isso.

 [Raphaella Reis](#) disse 4 meses, 3 semanas atrás:
Perdoem se chego atrasada, mas pelo que pesquisei no fórum, cabe a pergunta. Foi feito algum levantamento sobre acervos ou pesquisadores particulares que já estejam no caminho da digitalização de seus acervos e documentação de temas relacionados aos listados pelo Carlos Seabra no primeiro post, item 4? Se houve, há algum projeto de incentivo desse trabalho? Sei de alguns poucos sites que se dedicam a registrar fatos e pessoas importantes em nossa história e cultura na internet. Como incluí-los em iniciativas que visam fazer o que eles já fizeram? Há a possibilidade de dar prioridade a eles, ou subsidiá-los?

 [Camila Guimaraes Dantas](#) disse 4 meses, 3 semanas atrás:
Olá, mais uma pergunta neste território pouco mapeado: vocês conhecem alguma iniciativa brasileira para preservação de conteúdos nascidos digitais?

 [Alfredo Luiz Suppia](#) disse 4 meses, 3 semanas atrás:
Respondendo a Raphaella: eu não conheço nenhum levantamento mais sistemático sobre entidades com processo de digitalização em andamento. No campo específico do audiovisual, acho que a Cinemateca Brasileira dispõe de levantamento do tipo, porque tem captado acervos e realizado a digitalização. Conforme Carlos Seabra apontou, seria muito interessante a confecção de mapas de instituições com acervos em processo de digitalização ou "digitalizáveis". Entidades públicas forneceriam dados compulsoriamente, mas há um número razoável e importante de acervos particulares que precisam ser "garimpados". Esse trabalho poderia ser feito por um grupo especialmente designado, e o resultado seria de grande utilidade para orientação das iniciativas de digitalização e organização de plataformas colaborativas.

 [Alfredo Luiz Suppia](#) disse 4 meses, 3 semanas atrás:
Respondendo a Camila Dantas: não conheço tal iniciativa. O problema dos conteúdos nascidos digitais é a batalha dos formatos. Até onde tenho conhecimento (falo mais sobre o campo do audiovisual), os produtores de conteúdo digital cuidam da própria armazenagem, no formato que for mais conveniente. Ao contrário da película, que pode ser depositada na Cinemateca Brasileira, não conheço uma forma de depósito digital na instituição. Tal um aspecto bom para debate.

Fragmento de página do Fórum com a nossa primeira intervenção, em 31/07/2010. Acervo da autora.

No documento do Eixo, os arquivos nascidos digitais são citados de forma muito esporádica e referida ao campo do audiovisual, como na resposta de Alfredo Luiz Suppia à pergunta que fizemos no Fórum. Mas não há uma discussão direcionada para eles. E a Web, enquanto um possível alvo de preservação, não é sequer mencionada. Nos principais tópicos de discussão do texto (legislações, tecnologias, métodos, entre outros) o foco está sempre sobre a digitalização de acervos. O documento chega a utilizar o termo “memória digitalizada”, que nos parece de fato mais apropriado ao que está sendo tratado²⁰. A noção de “memória digital” é, assim, referida primordialmente ao patrimônio digitalizado.

²⁰ Trecho do documento em que aparece o termo “memória digitalizada”: “Interdisciplinaridade, interoperabilidade, interprogramação, intercomunicação e metadados, são termos mais comuns hoje em dia em diversas profissões que lidam com conteúdos digitais. Dessa perspectiva, a memória digitalizada de bens culturais, produções públicas e privadas, precisa diferenciar o uso que se faz desses conteúdos que, reunidos num suporte, o computador, tem adequações de usos, restrições jurídicas e caracterizações de propósitos diferentes” (LOUREÇO, 2009:s/p).

Observamos essa centralidade da questão da digitalização de acervos em outras instâncias de discussão e elaboração de políticas públicas. Além do Fórum e dos seus seminários presenciais, o Ministério da Cultura, em parceria com a Rede Nacional de Pesquisa (RNP) e com a Biblioteca Brasileira (USP), promoveu em 2010 o Simpósio Internacional de Políticas Públicas para Acervos Digitais. Os documentos que se originaram deste seminário (TADDEI, 2010; BRANT, 2010) demonstram que a questão dos arquivos nascidos digitais e da preservação da Web não era desconhecida. Alguns participantes internacionais a colocou em pauta. Mas os principais esforços – e isso aparece claramente nos corpo dos relatos – estavam direcionados para o problema da digitalização de acervos e a sua acessibilidade, incluindo aí uma discussão da legislação de direitos autorais.

Entre 2011 e 2010, a partir das articulações entre múltiplas instituições públicas e privadas responsáveis por acervos memoriais, que já vinha acontecendo nos inúmeros seminários, criou-se a Rede Memorial (rede nacional das instituições comprometidas com políticas de digitalização dos acervos memoriais do Brasil), com o objetivo de propor diretrizes para o processo de digitalização dos acervos e criar um fluxo contínuo de cooperação²¹.

Em meio a essa preocupação com a digitalização dos acervos, encontramos no discurso de algumas instituições públicas um uso do termo “memória digital” que, tal como aquele que identificamos no documento do eixo de Memória Digital, refere-se estritamente ao patrimônio digitalizado. É o caso da fala de Muniz Sodré, então presidente da Biblioteca Nacional, ao apresentar, em 2009, o projeto da Rede da Memória Virtual Brasileira:

Rede da Memória Virtual Brasileira é um projeto tanto ambicioso quanto inédito. Primeiro, a saudável ambição de automatizar e disponibilizar no ciberespaço os acervos de todas as instituições nacionais que disponham de um patrimônio visual ou textual. Se a preservação física de um bem simbólico já assegura a permanência intergeracional de um traço valioso numa cultura consolidada, acrescenta a promessa de ampliação do acesso a esse bem. É, por isto, um artefato poderoso de democratização do olhar e do conhecimento. (SODRÉ *apud* SUPPIA, DA SILVA, 2010:5)

21 A “Carta do Recife”, lançada por ocasião da criação da Rede Memorial, estabelece seis princípios ou compromissos assumidos pela Rede: Compromisso com acesso aberto (público e gratuito); compromisso com o compartilhamento das informações e da tecnologia; compromisso com a acessibilidade; padrões de captura e tratamento de imagens; padrões de metadados e de arquitetura da informação dos repositórios digitais; padrões e normas de preservação digital. No segundo Fórum, realizado em 2012, estes pontos foram ampliados e desdobrados em dez princípios. Disponível em: <http://redememorial.org.br/carta-do-recife-2-0/>. Acesso: 8/10/2013.

É interessante notar uma certa dissonância entre essa noção de memória digital – que permeava o discurso de instituições importantes como a Biblioteca Nacional e circulava numa rede de atores mais ampla como a do Fórum – e a noção de cultura digital, que norteava as políticas públicas do MinC e também se fazia presente no Fórum. Ambas compartilham certamente de uma mesma visão de cultura. Ambas também tomam a tecnologia digital como importante instrumento de disseminação das produções culturais em sua diversidade. Mas enquanto que a noção de cultura digital envolve um caráter produtivo da tecnologia digital, a noção de memória digital, focada no processo de digitalização, enfatiza um aspecto reprodutivo, uma vez que diz respeito preponderantemente em ações de salva-guarda do legado patrimonial.

Com esta análise não pretendemos, de nenhum modo, questionar a importância de uma política de digitalização dos acervos públicos, que pode de fato vir a contribuir para uma disponibilização mais ampla do patrimônio cultural. O que estamos buscando apontar é que a ênfase das políticas públicas na digitalização de acervos é parte de um modo de conceber a noção de memória digital que deixa de fora do seu escopo justamente aquilo que lhe seria mais específico – os arquivos nascidos digitais e, entre estes, os arquivos da Web.

Chegamos assim a um ponto de inflexão importante, que diz respeito à relação dos artefatos tecnológicos com as categorias sociais. Não foram poucas as vezes em que, ao longo da história, os objetos tecnológicos emergentes foram apreendidos a partir categorias já existentes. No caso em questão, embora uma nova categoria tenha sido nomeada – memória digital – o modo de conceituá-la não contemplava a sua possível especificidade/novidade. Tanto nas políticas públicas como nas discussões do Fórum, embora se visasse os usos das novas tecnologias, os debates giravam em torno de uma noção de preservação ancorada em pressupostos que datam de um paradigma moderno do patrimônio (GONÇALVES, 1999). Segundo esta perspectiva, aos bens patrimoniais correspondem objetos fixos, e não objetos dinâmicos tais como páginas da *web*, redes sociais, *tweets*, *streamings*, por exemplo.

É importante perceber que essa visão moderna do patrimônio, que se faz presente na maneira de definir a memória digital, destoa do modo como as políticas públicas concebem, de modo mais amplo, o patrimônio. Os posicionamentos do Ministério da Cultura e as discussões do Fórum estão em consonância com a ideia de que não só objetos estáticos e materiais devam ser preservados. Manifestações culturais de diversas ordens (dar exemplos) constituem um patrimônio imaterial que deveria receber igual atenção das políticas públicas. Assim, em paralelo à disparidade que há pouco apontamos entre as noções de “memória digital” e de “cultura digital”, observamos uma discrepância entre duas perspectivas distintas com relação ao patrimônio – a visão moderna, dirigida aos objetos digitais e uma visão ampliada, dirigida aos diversos bens culturais.

1.2.2 O silêncio sobre o passado da web no debate relacionado à memória digital no Brasil

A partir dessa incursão nos debates em torno do patrimônio digital no cenário nacional, entre o final da primeira década do século XXI e o início da segunda, com foco especialmente no Fórum de Cultura Digital, constatamos um vazio de discussão com relação à preservação da Web. Chamou nossa atenção o fato de num Fórum aberto de discussão sobre a cultura digital, realizado predominantemente na Web, não se considerar este “espaço” próprio no qual o Fórum se desenrolava como um possível campo de produção de memórias. A análise expôs também a complexidade das questões envolvendo o patrimônio digital: por um lado, uma perspectiva que podemos qualificar como ampla e aberta, no que concerne à cultura digital e ao patrimônio. Por outro, uma visão restritiva do que se nomeou como “memória digital”, predominantemente limitada ao processo de digitalização.

Temos elementos para supor que essa desatenção com relação aos arquivos nascidos digitais e à preservação da Web, não estava circunscrita à esfera governamental. Nota-se, por exemplo, a ausência de um posicionamento por parte de entidades como a Associação Nacional de História (ANPUH), que não participou do grupo de Memória Digital do Fórum. Percebemos também essa despreocupação com a preservação da Web nas entrevistas realizadas com empresários e pesquisadores da área de tecnologia voltados para empreendimentos na internet. Alguns deles demonstraram até mesmo uma reação de espanto a esta questão. Roberta Coelho, que participou na criação da ferramenta de busca Radix, afirmou em entrevista que para os profissionais envolvidos nesta área a preservação dos códigos ou mesmo das páginas de consulta não seriam tão relevantes, pois o foco está no esforço de se manter atualizado com as últimas tecnologias²²:

“Essa área da ciência da computação termina não sendo útil, ela se perde mesmo. Porque aquela tecnologia que foi útil dez anos atrás, você não vai ter uma razão para estudar ela hoje. Você sempre está querendo estudar a mais nova e sempre aquela correria que o conhecimento avança de uma forma que você não tem nem tempo de se manter atualizado. Então você terá menos tempo ainda para olhar o passado.” (COELHO, 2014)

Essa ausência de interesse pela preservação da Web no cenário brasileiro nos causou estranheza, impulsionando de certo modo a pesquisa. Nosso ponto de vista não se assemelha às posturas alarmistas. Mas se de fato podemos falar de uma cultura digital própria do nosso tempo, como não pensá-la como digna de preservação? O próprio Fórum de Cultura Digital, por exemplo, constitui, no nosso entender, um palco importante de debates político-culturais. Infe-

²² Roberta Coelho, entrevista em 24 de Junho de 2014. Anexo.

lizmente muitas das suas discussões não se encontram mais na rede. No decorrer da pesquisa, tentamos sem sucesso rever algumas das páginas da plataforma. Conseguimos acessar algumas delas unicamente porque as havíamos arquivado ou porque se encontravam disponíveis através do Internet Archive.

Entre os anos de 2013 e 2014, o cenário brasileiro não se modificou significativamente em relação à preservação dos arquivos nascidos digitais e da Web, exceto do ponto de vista normativo. Em abril de 2014, o Conselho Nacional de Arquivos (Conarq) lançou uma resolução²³ estabelecendo “Diretrizes para Implementação de Repositórios Digitais Confiáveis de Documentos Arquivísticos”. O documento apresenta um conjunto de critérios e normas para a criação e manutenção de acervos digitais, tendo em vista a sua permanência, confiabilidade e acessibilidade. Além de estabelecer critérios tecnológicos e de infraestrutura, a resolução abrange também aspectos políticos e procedimentos técnicos e administrativos necessários à implementação de repositórios digitais confiáveis (CONARQ, 2014:5).

Em termos de políticas públicas, observa-se um desdobramento das discussões e dos projetos voltados para a digitalização de acervos anteriormente esboçados. Em 2013, o Minc, junto com o Instituto Brasileira, a Petrobras e a Rede Memorial, criou o Edital Prêmio Memorial Digital, voltado para apoiar iniciativas de digitalização de acervos. Em 2014 foram selecionadas as dez primeiras instituições contempladas, que receberão infra-estrutura e treinamento de pessoal para realizar a digitalização de seus acervos. O edital constitui certamente a concretização de todo um processo de discussões e planejamento que se desenrola há pelo menos uma década. Portanto, não é de espantar que tendo em vista o vazio de discussão em relação aos arquivos nascidos digitais e à Web, no decorrer desse mesmo período, que hoje não tenhamos sequer esboçada uma política de preservação voltada para eles. Essa pesquisa navega neste campo pouco habitado institucionalmente, mas fértil em questões e possibilidades a serem trabalhadas.

²³ Resolução N.39, de 29 de abril de 2014.

2. MEMÓRIA DIGITAL: UM CAMPO DE MÚLTIPLAS CONEXÕES

O objetivo deste capítulo é apresentar alguns exemplos de experiências contemporâneas de utilização da web para elaboração, armazenamento e circulação de memórias ou testemunhos digitais. Ao percorrer essas experiências, numa espécie de etnografia digital, interessa-nos igualmente refletir sobre os termos do debate implicados na noção de criptografias de memória, que perpassa esta pesquisa. Este termo talvez pudesse ser facilmente substituído, por exemplo, por tecnologias da memória, um conceito adotado por Marita Sturken (1996). Tal conceito põe a ênfase das análises nas práticas e operações da memória, reforçando o aspecto transiente e conflituoso. Para entender essas relações e esboçar nossa apropriação deste conceito na noção de criptografias, que se refere às práticas de memória relativas ao arquivamento da web, faremos um percurso pelos principais nós na constituição de área de saber sobre a memória.

Na década de 1990 houve uma explosão de memórias sociais que se manifestaram de várias formas, através de exposições, filmes, projetos testemunhais. Denominado de “boom de memórias” (NORA, 1993) ou musealização (HUYSSSEN, 2001), tal processo tem sido amplamente estudado e está associado ao surgimento de uma área interdisciplinar, dedicada ao estudo sobre a temática da memória. A criação do periódico *Memory Studies*, pela Sage em 2008, aponta na leitura de Marita Sturken (2008:3) para um momento de institucionalização deste campo de pesquisas que embora incorpore muitos problemas e metodologias da História, lança mão de um enfoque mais próximo a chamada área dos Estudos Culturais. Esta pesquisa, elaborada no âmbito do Programa de Pós-Graduação em Memória Social constitui-se igualmente um reflexo de como este trabalho insere-se justamente nesta área de conhecimento, este capítulo tratará em seu primeiro item de aspectos teóricos relevantes para se pensar as práticas de memória na contemporaneidade.

Para abordar estas questões adotamos a estratégia de explorar as possibilidades teóricas apresentando dois casos que remetem a aspectos distintos. A questão da criação de plataformas, que propiciam a criação e circulação de testemunhos nascidos digitais, é o ponto convergente entre os casos, que distinguem-se bastante em outros aspectos. A intenção ao escolher projetos memorialistas de natureza distintas foi justamente a de poder através deles apontar as questões em pauta no campo interdisciplinar da memória. Ao mesmo tempo é intenção deste capítulo, mostrar como a noção de patrimônio digital está se conformando no cenário internacional, em especial na Inglaterra. O exemplo descrito no terceiro item deste capítulo torna-se, desse modo, um contraponto para o cenário brasileiro já explorado no primeiro capítulo.

2.1 Enredamentos teóricos no campo de estudo das memórias sociais

No polêmico campo de saber da memória social este é um tema que tem recebido uma maior atenção, porém ainda são relativamente escassas as elaborações teóricas que tratem especificamente das relações entre as práticas memorialistas e as novas tecnologias da informação. Neste item nos propomos definir os termos utilizados neste trabalho, inserindo-o em um debate em pleno curso, mas que remete a um longo percurso de desenvolvimento da sociologia da memória.

A interdisciplinaridade deste campo de pesquisa não pode ser negligenciada, pois é na difícil articulação de conceitos e dados que uma pesquisa se constitui. De acordo com Regina Abreu:

“A multiplicidade de caminhos na relação entre teoria e pesquisa é ainda mais difícil no âmbito da memória social, uma vez que não estamos diante de um campo disciplinar. A memória social é um campo de estudos visitados por pesquisadores das mais diversas procedências e que fazem uso de múltiplos referenciais teóricos” (2005:30)

A vereda que abriremos aqui busca a compreensão de alguns nós importantes na constituição desta área de conhecimento. Não se trata de um levantamento exaustivo, mas sim da apreciação de forma sucinta de alguns marcos teóricos importantes. O estudo dos grupos e da coesão social apresenta-se como um ponto central, desde o início do século XX, nas pesquisas sobre a constituição da memória social. É portanto indiscutível a importância desse modo qualquer estudo que se proponha a obra de Maurice Halbwachs (1877-1945). A centralidade da obra do sociólogo francês é ponto pacífico, embora as controvérsias e atualizações multipliquem-se no âmbito das mais recentes pesquisas eja para opor-se a ela ou para reconhecer-se como parte de uma linha investigativa devedora dos estudos inaugurais desenvolvidos pelo sociólogo francês (Klein, 2001; Gondar, 2005).

Nesse sentido, vale mencionar que a obrigatória referência aos estudos desenvolvidos por Halbwachs não corresponde a existência de um consenso sobre a nomenclatura ou interpretações, nem mesmo entre aqueles que se dizem seus herdeiros. A apreensão da memória a partir de um enfoque social, alicerçado na sociologia de E. Durkheim (1858-1917) criou toda uma nova ordem de questões relacionadas aos atos de lembrar e esquecer como práticas sociais. Além disso o diálogo proposto com a filosofia, principalmente Bergson, e é um ponto nevrágico do construindo uma abordagem alicerçada nos pressupostos da sociologia de Durkheim, o importante do pensamento de Maurice Halbwachs constitui um marco na apreensão da memória a partir de um enfoque social. Sua importância situa-se justamente na criação de uma nova ordem

de questões relacionadas à memória, que deixava de ser perscrutada unicamente como um atributo individual (Sepúlveda, 2003, Hutton, 1993).

A memória é coletiva, nos termos de Halbwachs, na medida em que seria constituída por imagens e quadros esquemas; a incorporação do passado é um fator de coesão dos grupos. Os indivíduos não recordam sozinhos é a partir dos quadros sociais que a memória deveria ser investigada. O termo “quadros sociais”, advindo dos estudos literários, era no período bastante utilizado pela crítica literária com relaciona-se a o significado de um sistema de referências temporais, mas na obra de Halbwachs há uma semelhança semântica, porém a noção é ampliadamente ampla. O autor associa quadros sociais a um “sistema de representações” que podem ser sistemas de lógica, de cronologia ou topografia, que antecipam a lembrança fornecendo assim um sistema global de localização do passado no presente (NAMER, 1994: 325).

A sociologia da memória tornou-se, sobretudo nas últimas décadas do século passado, um campo fundamental de pesquisas. E, para além das fronteiras da Sociologia, a memória tornou-se um tema tão relevante que uma extensa terminologia foi sendo criada para tratar dos fenômenos em curso: “abusos da memória” (TODOROV, 2002), “musealização” (HUYSSSEN, 2001), “boom de memórias” (NORA, 1993). A disciplina histórica tomou-a como objeto, a partir das inflexões próprias de uma historiografia que abandonava os pressupostos historicistas que pressupunham uma lógica dicotômica entre a Memória e a História (HUTTON, 1993).

O historiadores entraram no debate no sentido de distinguir as narrativas do passado elaboradas de modo profissional, seguindo determinadas regras metodológicas da disciplina histórica, dos relatos de caráter memorialistas (LE GOFF, 1998). Ao mesmo tempo a História do Tempo Presente incorpora essa perspectiva de análise, que procura dar legitimidade à produção do conhecimento profissional, como afirma Marieta de Moraes Ferreira:

“Os historiadores das décadas de 1980-1990 enfrentam a necessidade — recorrente na história da disciplina — de defender a autonomia da história e, portanto, de reafirmar sua intenção contra a redução da história à ficção, às falsificações negacionistas, às derivas memoriais, às instrumentalizações sociais e políticas da história. (...) Diante do crescente e permanente interesse no presente, tanto pela comunidade de historiadores quanto por diferentes grupos sociais, e da afirmação desse campo de trabalho, a história do tempo presente tem buscado superar esses questionamentos para se legitimar mapeando melhor seus limites, suas especificidades. Em busca de uma definição mais precisa de seu objeto, de suas metodologias, abre caminhos para novas investigações tais como a história das memórias coletivas e os usos do passado e as comemorações.” (2012:117)

Neste mesmo sentido é que podemos compreender a crítica elaborada por Beatriz Sarlo (2005) frente a emergência de uma cultura permeada pela prevalência do relato testemunhal. A guinada subjetiva estaria justamente no pressuposto de que “não há verdade, mas paradoxalmente, os sujeitos tornaram-se cognoscíveis” (2005:39). Percebe-se assim, um entremear-se entre o campo dos estudos culturais, a sociologia da memória e a historiografia contemporânea²⁴.

Paralelamente, o arquivo deixa de ser visto como um mero depósito de documentos e passa a ser investigado como elo fundamental entre o passado e o presente. Tal abordagem inaugura-se com Michel Foucault que pensa o arquivo a partir de uma lógica disciplinar, tomando-o muitas vezes como metáfora e não apenas como entidade física (DELANDA, 2005; HEYMANN, 2009). Perguntar sobre o modo de organizar as informações em um arquivo é tão importante quanto o conteúdo de um documento específico. O arquivo emerge como um objeto de estudo relevante na apreensão das relações de poder e os regimes de verdade, seguindo o vocabulário de Foucault. No próximo capítulo, será realizada uma discussão sobre a metáfora do arquivo assim como as possibilidades de dimensões diversas ou as implicações do arquivo digital como gerador de uma memória em curto circuito (ERNEST, 2013). Neste momento trata-se apenas de apontar a extrema importância que o arquivo, como objeto e metáfora passam a exercer neste campo (MANOFF, 2010).

Nesse sentido, consideramos que o arquivo pode ser entendido como uma tecnologia da memória, mas não apenas como manancial documental. Neste trabalho ao mencionarmos os arquivos nascidos digitais estamos tomando o termo de uma maneira bem menos abrangente. As tecnologias da memória devem ser apreendidas em seus contextos a fim de revelarem as práticas a elas associadas. Assim, ao nos debruçarmos sobre os imbricamentos entre a memória e os arquivos nascidos digitais, estaremos preocupados em descrever e refletir sobre as práticas de memória, tal qual descreve Marita Sturken:

“the concept of memory practices allows for an emphasis on the politics of memory, precisely because of the ways in which the production and construction of memory through cultural practices has as its foundation the notion that memories are part of a larger process of cultural negotiation. This defines memories as narratives, as fluid and mediated cultural and personal traces of the past. A practice of memory is an activity that

24 A seguinte afirmação de Marita Sturken é elucidativa sobre as dinâmicas entre essas áreas de estudo: “it is interesting to note for which disciplines the topic of memory is considered to be a problem, and for which is it merely an area of expansion. For instance, while the study of memory has been subject to concern within the field of history, for its potential to be amorphous and undisciplined as an area of study and for its implicit critique of historical studies, its emergence as a topic in literature or media studies has been relatively uncontroversial. Yet, it is the case today that much of the most interesting work pursued in memory comes from historians or is deeply indebted to the innovations of historiography over the past few decades.(2008:2)

engages with, produces, reproduces and invests meaning in memories, whether personal, cultural or collective. (STURKEN, 2008:3)

Explorar as mediações entre as práticas e as materialidades digitais é objetivo dos próximos itens, onde os objetos escolhidos evocam práticas de construção das memórias individuais e sociais. O balanço proposto por Sturken (2008) sugere que a questão da terminologia pode ser cotejada de modo produtivo. A questão não seria optar estritamente por referir-se aos usos do passado como memória social ou memória cultural, mas sobretudo explicitar a perspectiva adotada na investigação do tema, em cada trabalho e contexto específico.

A mediação exercida pelas tecnologias no campo da memória social é um campo que vem sendo explorado por vários ângulos. A noção de criptografias da memória apenas acrescenta algo na discussão se vista dentro do contexto aqui elaborado, e o modo como está diretamente associada a noção de memória socialmente constituída. As evidências coletadas ao abordar formas distintas de registro e circulação das memórias sociais na web que apresentaremos a seguir buscam ilustrar a diversidade constitutiva do registro digital, cujo elemento básico é o website. As criptografias, no sentido de mediações exercidas pelas tecnologias (Van Dick, 2007) nos servem aqui para ajudar na investigação de aspectos culturais presentes também na estruturação das plataformas digitais para coleta de dados, como veremos no próximo item.

2.2 Memórias anônimas na web²⁵

Nas linhas que se seguem nos propomos o seguinte percurso: pensar os registros digitais através de dois projetos (*Postsecret* e *Memoryshare*), investigando de que modo tais objetos podem colocar questões para o modelo teórico sobre a memória social proposto pioneiramente na obra de Maurice Halbwachs (1950). Nossa proposta neste item é tratar alguns problemas conceituais da contemporaneidade a partir do enigma das memórias anônimas e apontar algumas possibilidades teóricas para abordar tal fenômeno. A seguir faremos uma descrição dos objetos selecionados, já tecendo nossa argumentação que tomará forma no segundo momento.

²⁵ Este item é uma versão ligeiramente modificada do artigo *Memória Anônimas* (DANTAS, DODEBEI) apresentado no IV Congresso Online del Observatorio para la Cibersociedade, em 10 de Novembro de 2010. ISBN 978-84-613-7299-7.

2.2.1 *Memoryshare* ou é possível compartilhar memórias via internet?

O projeto *MemoryShare*, criado em julho de 2007 pela BBC de Londres, tem por objetivo o registro de memórias a partir de 1900 até os dias de hoje. Os usuários da plataforma digital podem compor suas reminiscências não apenas com palavras, mas também com fotos e vídeos. Cada integrante do projeto possui um cadastro e pode inserir suas memórias, apagar as anteriores ou editá-las a qualquer momento. Há registros autorais, ou seja, que contém nome completo do autor; porém, há também uma grande quantidade de registros anônimos ou com assinaturas que serão reconhecidas apenas por um determinado grupo. Importa notar que a autoria não é obrigatória para inclusão do relato ou foto no acervo digital. Esse é um aspecto bastante interessante deste arquivo que transpõe para o universo dos registros de memória um aspecto comum em redes sociais na web, nas quais não é obrigatória uma identificação completa.

A partir de uma navegação impressionista no acervo *Memoryshare* podemos dizer que se trata de um artefato de caráter bastante fragmentário e que atende às demandas da BBC para criação de conteúdos para documentários e programas de rádio e TV. De fato, há uma vasta gama de temas ali registrados: casamentos, eventos públicos, guerras, encontros, etc. Todos estes registros inserem-se em um software de visualização que apresenta cada “memória” como um ponto numa linha do tempo em espiral, compondo uma imagem psicodélica e proporcionando uma escolha dos itens de uma maneira bastante diversa daquela de um leitor de uma lista em ordem alfabética, por exemplo. Descrever este sistema de apresentação dos registros é um esforço de transpor em palavras este tipo de navegação que, no entanto, é uma das características mais importantes do *website*. Numa tentativa de apreensão deste objeto, optamos por descrever três interfaces. A primeira delas é justamente esta imagem colorida e abstrata da abertura que é a via de acesso ao conteúdo do website.

Imagem 11 - Página inicial do website *Memoryshare*



Tela do portal BBC Memoryshare. Fonte: Screenshot da coleção pessoal da autora.

Cada grão colorido que integra a espiral corresponde a um registro. A cor deste grão é gerada a partir do dia da semana da memória em foco e sua forma deriva da primeira letra do registro que é utilizada numa composição abstrata. Vale notar que não é uma imagem estática: é o movimento que nos guia pela torrente de memórias. Para ler um destes registros é preciso tão somente clicar no item e aportar numa interface bem mais convencional, com título e caixa de texto. Essa será nossa segunda parada.

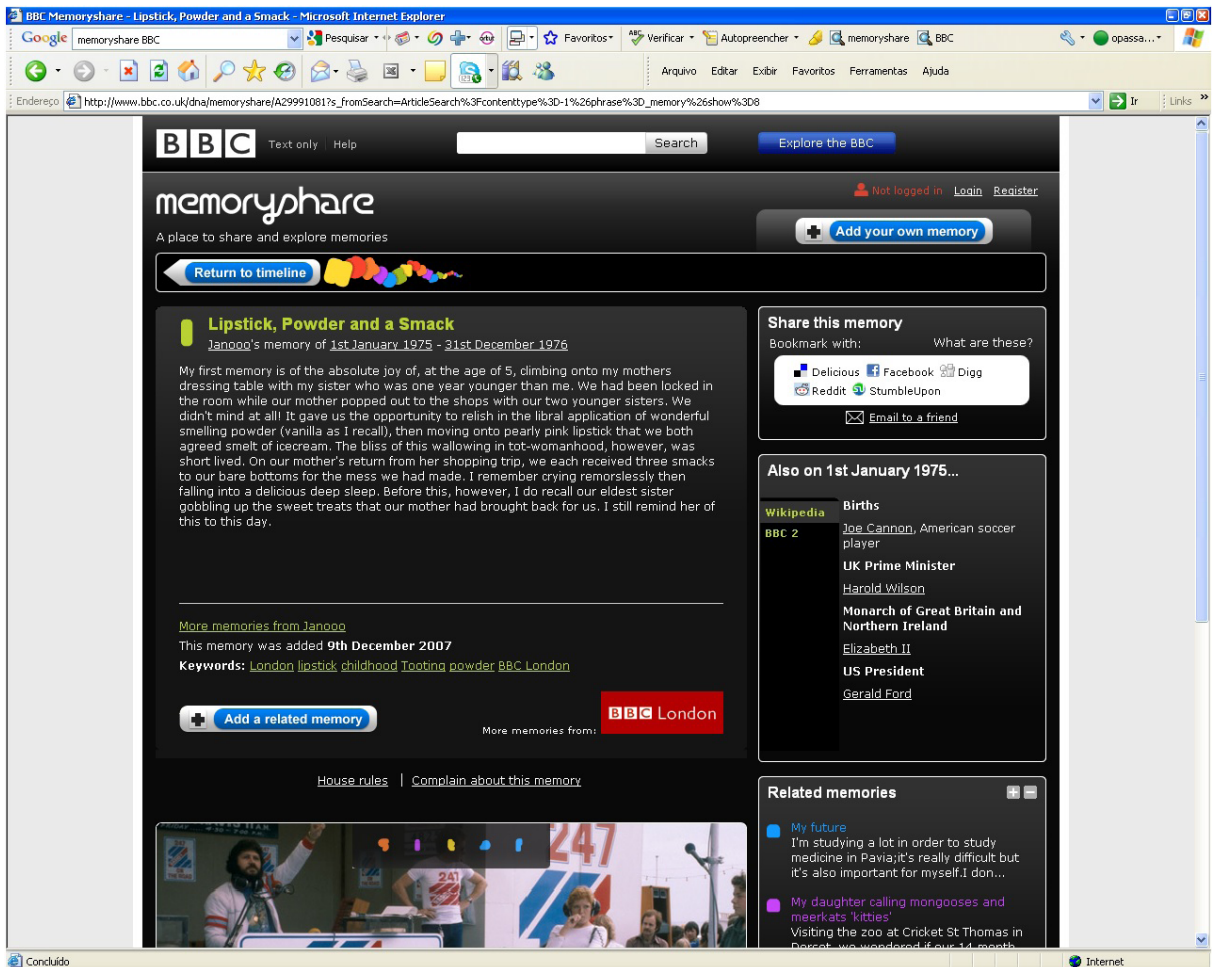
Imagem 12 - Selecionado um registro no website *Memoryshare*



Tela “The Memories of Today”. Fonte: Screenshot da coleção pessoal da autora.

A interface das memórias propriamente ditas pode ser descrita como uma caixa de texto contendo ao final *tags* e *links* para o acervo. Em nossa pesquisa, de caráter impressionista, pudemos verificar uma fragmentação das narrativas, muitas delas referentes a fatos da vida cotidiana. Títulos como “batom, pó e palmada” nos indicam o caráter prosaico de alguns relatos.

Imagem 13 - Acesso à memória “batom, pó e palmada” no website *Memoryshare*



Tela: Lipstick, Powder and a Smack. Fonte: Screenshot da coleção pessoal da autora.

O da imagem acima, por exemplo, conta a primeira lembrança de sua autora identificada apenas por Janoo, ocorrida em janeiro de 1976. Não sabemos ao certo a identidade de Janoo, sabemos, tão somente, que num momento longínquo de sua infância ela usou a maquiagem de sua mãe e, por isso, levou algumas palmadas. Sabemos também a partir deste relato que narrar tal lembrança do cotidiano é algo compartilhado pelos sujeitos que habitam o ciberespaço. Janoo nos coloca na fronteira do público e do privado: entre o relato autoral e o registro anônimo.

Cabe perguntar: qual é a interface de construção de tais relatos? Eis nossa última parada. Para inserir um relato o usuário deve primeiramente cadastrar-se preenchendo um formulário simples no qual não é exigida a inserção de dados pessoais (nem mesmo o nome). O que importa é ter uma conta de e-mail válida, este sim é o passaporte. Depois desse primeiro passo é que o usuário alcança a interface de inclusão de relatos que possui um enquadramento bastante tradicional. É um módulo composto por quatro perguntas:

1. O que aconteceu?
2. Quando?
3. Como?
4. Onde?

No interior de cada módulo há caixas com limite de espaço e no caso dos referenciais temporais é dada opção para inserção de períodos. Ao final, pede-se que o usuário selecione palavras chaves que serviram de categorias do relato. O website conta com filtros automáticos (restringindo o uso de determinados termos) e permite a inclusão de relatos apenas em língua inglesa.

A partir deste breve percurso, podemos perceber alguns aspectos desta plataforma de registros digitais. A fragmentação dos temas e a ausência de qualquer esforço reflexivo sobre as memórias ali registradas, parece nos colocar diante daquilo que Beatriz Sarlo (2007:13) chamou de *uma cultura de memória ausente de pensamento sobre a memória*. Por outro lado, onde encontramos alguma elaboração encontramos uma proposta de criar uma “história dos cidadãos” como lemos no trecho abaixo retirado do editorial do site:

“Memoryshare is about putting your memories online to build up a unique picture of living in the British isles since 1900.

It’s unique because Memoryshare is about building up a picture of our history from Britain’s citizens. Anyone who lives, learns, works or plays in this country can contribute to recording what life was like living in the 20th century and continuing into the 21st.” (BBC, 2013)²⁶

De alguma forma, esta passagem nos revela um “desejo de memória” associado a um projeto nacional que se quer perpetuar para as futuras gerações. Como entender tal projeto amalgamado em uma plataforma de visualização tão fragmentária? Como pensar a construção de identidade a partir de registros anônimos? Eis algumas questões colocadas por este artefato cultural, o portal Memoryshare.

26 Tradução livre: “Com Memoryshare você coloca suas memórias online para que se possa construir um retrato da vida nas Ilhas Britânicas desde 1900. É um projeto único que visa construir um retrato da história britânica a partir de seus cidadãos. Qualquer pessoa que tenha vivido, estudado, trabalhado ou se divertido neste país pode contribuir registrando como era viver a vida no século XX e no século XXI.” Ver, *What is memoryshare?*, Disponível em: http://www.bbc.co.uk/humber/content/articles/2008/01/15/02_memory_share_what_feature.shtml (acessado em 15 de julho de 2015).

2.2.2 PostSecret - o que são os segredos públicos da nossa rede digital?

PostSecret é um *website* criado em novembro de 2004 por Frank Warren, um artista norte-americano, cujo objetivo é publicar postais anônimos que contêm um segredo. De início, Frank Warren distribuiu aproximadamente três mil postais já selados nas ruas de Washington (DC) para que as pessoas pudessem enviar pelo correio e assim participar de seu projeto. Este foi o ponto de partida do projeto que tornou-se um dos sites mais visitados da web norte-americana. Hoje a comunidade virtual *PostSecret* já possui versões em espanhol, francês, alemão e coreano. Desde então, Warren recebeu aproximadamente 175 mil cartões postais e publicou apenas uma porcentagem dos mesmos. Recentemente foi criado um outro website, denominado *PostSecret Archive* para publicar uma parte dos postais não escolhidos ao longo dos anos (WARREN, 2009).

Imagem 14 - Imagem e conteúdo de um postal publicado no website *PostSecret*

The screenshot shows a web browser window displaying the PostSecret Archive website. The main content is a handwritten postcard with a circular stamp. The text on the postcard reads:

At my best friends graduation today I heard a fire-engine and could only think of a bomb going off, the F.D. responding. Firecrackers went off and I heard machinagun fire. When I drive in my car, as I look for oncoming traffic I also check local buildings and houses for snipers.

I have been out of the military for 2 years and was only in Iraq for 4 months. I am so fucking scared, of everything.

Below the postcard, the original message is displayed:

-----Original Message-----
Sent: Sunday, May 28, 2006 10:33 PM

To the person who was in Iraq for 4 months. I was in Iraq for 13 and I hate fireworks now too. I don't like the 4th of July. I almost puked at last years fireworks. Twice.

I know how you feel. It sucks. I would never wish it on my worst

The website interface includes a navigation menu on the right with a list of months from July 2009 to January 2005. There are also advertisements for 'SecretZen' and 'Books We'.

O número significativo de postais recebidos pelo projeto nos coloca diante de uma das questões fundamentais sobre a internet: a exposição da vida privada. Ou melhor, como destacam alguns autores, o que está em jogo é uma redefinição do público e do privado (VAN DIJK, 2007:53-66). Os blogs são relatos de escrita íntima, porém diferentemente dos diários do século XIX, são escritos para serem imediatamente consumidos. Este é o caso do *PostSecret*, que diferentemente do *MemoryShare* não possui um discurso de construção de uma memória, mas nos aponta para a questão da formação de um acervo de “pequenos testemunhos” selecionados e de caráter anônimo. Ao mesmo tempo, trata-se de um projeto artístico de interação com o público que denota uma certa possibilidade de narração do cotidiano. Trouxemos este exemplo como contraponto por ser algo significativo na cultura da rede, citado por muitos blogs, eleito em 2006 como melhor blog norteamericano, e publicado em uma série de livros.

Frank Warren, criador do projeto, descreve com as seguintes palavras a acolhida a *PostSecret*:

“I stopped passing out postcards but secrets kept coming. Housemade postcards made from cardboard, old photographs, wedding invitations and other personal items artfully decorated arrived from all over the world. Some of the secrets were written in Portuguese, French, German, Hebrew and even Braille. One of the first *PostSecret* I received looked like nothing more than a worn postcard filled with two shopping lists. But squeezed into the corner was a soulful admission: I’m still struggling with what I’ve become. (...) Some of the most beautiful postcards in this collection came from very painful feelings and memories.” (WARREN, 2005)²⁷

Embora *PostSecret* não seja a rigor um projeto de memória, o arquivamento de tais mensagens nos coloca diante de um emaranhado de questões sobre as práticas contemporâneas de viver, narrar e arquivar o cotidiano. O anonimato está ali posto como condição de enunciação e assim será preservado. Como pensar a construção de um acervo de relatos individuais e anônimos? Que categorias podem ser manejadas para pensar tal elaboração? É possível pensar na categoria de grupo ou estamos diante de um artefato cultural que embora coletivo não pressupõe laços sociais entre os seus membros? E mais: como pensar a circulação da tecnologia ali envolvida que vai do velho cartão postal que é digitalizado e transposto para web? Essas são

27 Warren, F. (2005). *PostSecret: Extraordinary confessions from ordinary lives*. <http://postsecret.blogspot.com> (acessado em 20 de julho de 2010) Tradução livre: “Eu parei de distribuir cartões postais, mas eles continuaram chegando. Cartões caseiros, feitos de papelão, com fotografias antigas, convites de casamento e outros itens de caráter pessoal que eram artisticamente decorados, chegavam de todas as partes do mundo. Alguns dos segredos eram escritos em português, francês, alemão, hebreu e até braile. Um dos primeiros segredos que recebi parecia nada mais que um cartão usado preenchido com duas listas de compras. Mas, espremido entre elas havia esta profunda confissão: Eu ainda estou lutando com o que me tornei.”

algumas das questões que podem ser pensadas a partir de uma breve análise deste artefato que torna as fronteiras entre arte, registro e memória bastante difusas.

2.2.3 Anonimato e registro: como pensar o coletivo?

A proposta deste item é relacionar o conceito de memória social com a dissolução das fronteiras entre o público e o privado que vivenciamos na contemporaneidade (SIBILA, 2001). Em recente balanço teórico sobre o tema, o sociólogo norteamericano Jeffrey K. Olick propõe abandonar o termo “memória coletiva” em favor da expressão “sociologia das práticas e produtos mnemônicos” que agregaria os *insights* mais relevantes da obra de Halbwachs. Para Olick: “To focus on collective memory as a variety of products and practices is thus to reframe the antagonism between individualist and collectivist approaches to memory more productively as a matter of moment in a dynamic process” (2007:58)²⁸. A noção central adotada é de que a memória é simultaneamente individual e coletiva, tratando-se de um fenômeno que se expressa de múltiplas formas e em mídias diversas. Mas, então, como apreender tal fenômeno? Olick aponta a necessidade da questão entre indivíduo e sociedade, especialmente no que se refere às práticas relacionadas à memória, devem ser perscrutadas teoricamente, mas em conjunção com pesquisas empíricas. Ou seja, para o pesquisador não seria possível advogar uma definição a priori de tais relações, propondo que discernir entre individual e coletivo é uma tarefa de todos os estudos empíricos sobre o tema que procuram desvelar as nuances possíveis em determinados objetos.

Do mesmo modo, o modelo conceitual proposto por José van Dijk (2007) de uma memória mediada, ao mesmo tempo social e individual, pode ser uma possibilidade analítica para enfrentar as modalidades contemporâneas de estabelecimentos de vínculos entre o passado e o presente. A pesquisadora holandesa elege como objeto de estudo registros de caráter marcadamente individual como os álbuns de fotografias digitais ou vídeos caseiros, tratando-os como objetos de análise cultural. A tecnologia e sua relação com a constituição de uma ecologia cognitiva (LEVY, 1998) são incorporados como problemas centrais na argumentação da autora, que oferece um rico panorama teórico sobre o tema entrelaçado a análises empíricas relevantes. O seguinte trecho elucidada sua interessante proposta:

28 Tradução livre: “Focalizar a memória coletiva como uma variedade de produtos e práticas é repensar o antagonismo entre as perspectivas individualistas e coletivistas em prol de uma abordagem mais produtiva que trate a memória como um problema do momento, inserida em um processo dinâmico”.

Mediated memories are the activities and objects we produce and appropriate by means of media technologies, for creating and re-creating a sense of past, present and future of ourselves in relation to others. Mediated memory objects and acts are crucial sites for negotiating the relationship between self and culture at a large, between what counts as private and what as public, and how individuality relates to collectivity. (...) Mediated memories refers both to acts of memory (construing relational identity etched in dimension of time) and to memory products (personal memory objects as sites where individual minds and collective culture meets)²⁹. (2007:21)

O pressuposto de José van Dijk é o de que os artefatos individuais estão inseridos em modos compartilhados socialmente de registrar e lembrar o passado. Tais modos estão, por sua vez, estreitamente relacionados às mídias utilizadas. Fotografias ou imagens em movimentos do núcleo familiar evidenciam uma prática compartilhada de viver o presente. Tal como J. Olick, a autora reconhece a importância de M. Halbwachs para a fundamentação de um campo de saber sobre a memória, porém distintamente do pesquisador norte-americano, José van Dijk propõe um modelo teórico que transcende o pai fundador da sociologia da memória. As obras de H. Bergson e G. Deleuze constituem-se em pontos importantes da sua abordagem que procura estabelecer relações com o campo da psicologia cognitiva. A partir deste diálogo teórico emerge um modelo que trata também da possibilidade criativa nos registros da memória, mas que se propõe a uma investigação de caráter cultural ampliando as premissas de Bergson e Deleuze (2007:27-9). As aproximações teóricas e as análises empíricas propostas em *Mediated Memory in Digital Age* (2007) fazem deste trabalho uma referência importante para aqueles que pretendem investigar as conexões entre as elaborações da memória no mundo contemporâneo.

A partir das trilhas conceituais abordadas é que nos propomos a tecer algumas reflexões sobre os artefatos digitais escolhidos, *Memoryshare* e *Postsecret*, que nos colocam diante de um modo específico de relação com o passado e, ao mesmo tempo, com as diferentes possibilidades teóricas de pensar tais práticas. A relação anônima e fragmentada com o vivido é um fenômeno comum nos dois objetos culturais, muito embora haja distinções entre os mesmos.

29 Tradução livre: “Memórias mediadas são as atividades e objetos que produzimos e nos apropriamos utilizando as tecnologias da informação, para criar e recriar um senso de passado, presente e futuro de nós mesmos em relação com os outros. Objetos e atos de memórias mediadas são cruciais na negociação da relação entre o eu e a cultura, entre o privado e o público, e como a individualidade relaciona-se com a coletividade. (...) memórias mediadas refere-se a ambos: atos de memória (constroem uma identidade relacional situada numa dimensão temporal) e produtos de memória (objetos pessoais de memória onde a mente individual se entrelaça com uma cultura coletiva).”

No caso do projeto da BBC o anonimato não é obrigatório, mas uma navegação impressionista nos mostrou uma grande quantidade de memórias anônimas ou com codinomes compartilhados em grupos restritos. Tal observação nos coloca diante da necessidade de refutar uma associação automática entre memória e identidade. Estaríamos então diante de um novo paradigma da memória social? De que maneira tais artefatos digitais, tais mediações da memória, poderiam evidenciar tal virada?

É preciso redobrar nossa cautela ao embarcarmos em tais reflexões. Em primeiro lugar, é bom ressaltar que nossos objetos estão mergulhados em um universo múltiplo de conexões, entre as mídias e entre outras formas de registros do passado, como a oralidade ou os monumentos públicos. O esforço de pesquisar e pensar sobre os problemas trazidos por tais objetos nas novas mídias não exclui todo um universo paralelo onde as questões de identidade e de grupo são fundamentais para compreender as mais diversas práticas memorialísticas. Nesse sentido, o primeiro ponto relevante a apontar é que tais objetos só podem ser plenamente compreendidos a partir de uma investigação minuciosa que integre sua relação com outras mídias e com as práticas sociais mais amplas. No caso do projeto *Memoryshare* seria fundamental cotejar os usos dados pela BBC ao conteúdo ali gerado e, ao mesmo tempo, relacionar o estilo das contribuições com outras formas de relato em circulação, como o Weblog.

A partir de nossas considerações acima, redobrando o cuidado com generalizações apressadas, pensamos poder apontar *três pontos cardeais* para a análise de tais artefatos de memória em suporte digital. Apontamos aqui alguns caminhos e relações entre concepções teóricas e os objetos digitais.

Primeiro: as escritas e a tela. As interfaces do passado em plataforma digitais refletem e criam as novas formas de registrar e de ler na tela do computador. Os estudos sobre as novas práticas de leitura nos mostram o quanto os percursos mutáveis e a velocidade de apreensão de um texto tornaram-se variáveis fundamentais na contemporaneidade. Assim, ao tratar dos fragmentos do passado encontrados no *Memoryshare* é preciso situá-los neste contexto de letramento digital. Investigar como vem se processando na rede mundial de computadores a construção das memórias sociais, em milhares de redes, em imensos agregados informacionais, em novos e distintos acervos digitais, é pois tratar deste universo de novas práticas de registro. Tais práticas estão associadas a uma produção e a uma recepção distinta utilizando os termos de Roger Chartier, historiador das práticas de leitura que possui uma obra fundamental sobre o tema (2000). Assim, ao pensar sobre os cartões postais de PostSecret devemos situá-los dentre o universo cada vez mais curto e fragmentário dos registros online. Do blog ao microblo-

guing, assistimos hoje ao encurtamento das sentenças e a incorporação das imagens na lógica da comunicação interpessoal. A estética da colagem está presente nos mais diversos artefatos culturais e emerge a partir da confluência de fatores culturais e tecnológicos; emerge na própria delimitação das mediações (VAN DIJK, 2007:149-50).

Segundo: memórias do imediato. Tratando-se de objetos culturais situados de natureza fluida a partir de que categoria poderíamos propor uma análise? O conceito de uma memória de distribuição e elaboração imediata (“memory on-the-fly”), proposto por Andrew Hoskins (2009), nos traz importantes possibilidades analíticas. O sociólogo inglês cria tal conceito a partir de uma categoria de criação de programação de informática; “on-the-fly” é um modo de programar softwares à medida em que eles estão em funcionamento. Ao trazer esta imagem, estreitamente associada ao mundo digital, Hoskins aponta para a necessidade de trazer perspectivas de análise próprias a este meio de difusão e criação das memórias sociais. De acordo com Hoskins: “the array of technological advances that have transformed the temporality, spatiality, and indeed the ‘mobility’ of memories to an extent that even the dynamics of the emergent field of memory studies seem unable to keep pace with what I propose here is a new ‘distribution’ of memory. That is to say, the very condition of remembering in our new media ecology is that which is actively and re-actively constructed ‘on-the-fly’, notably through a range of everyday digital media”. (2009:67)

Terceiro: memórias e mídias. As interfaces entre a tela e os indivíduos são também perpassadas por outras janelas midiáticas. Seguindo a reflexão de Peter Burke, sobre a história social da mídia, pensamos ser fundamental situar cada mídia em uma temporalidade, explicitando que as novas mídias não excluem sistemas anteriores. Articulado essa ideia geral com o campo da memória, pensamos ser fundamental a intenção de situar os objetos em relação a outras produções. Por exemplo, seria factível pensar numa aproximação visual entre o caleidoscópio de pontos, que se constitui a forma de categorizar os registros no portal *Memoryshare*, e a representação cinematográfica dos circuitos de memória do filme *Eternal Sunshine of a Spotless Mind* quando os funcionários da Lacuna procuram apagar os vestígios de Clementine, vê-se cenas aleatórias numa espécie de tornado em espiral com muitas cores e abstrações. De que maneira ambas as representações remetem e recriam pressupostos comuns sobre a memória individual? Esta seria uma via possível de alargar o entendimento da plataforma *Memoryshare* onde o aspecto visual não deve ser menosprezado. Seguimos, neste ponto, os encaminhamentos propostos por Martin Zierold para o estudo sobre memória e mídia: “Considering the various contrary tendencies and the trend towards a pluralization of process of remembrances in society, the connection between the media and memory cannot be described only in theoretical

manner, but must also be observed in detail empirically.” (2007:406)³⁰. Zierold acredita que estudos comparativos e de caráter empírico podem ajudar a trazer novas nuances teóricas para o estudo das mais recentes formas de registro e apropriação do passado.

Esta é também a nossa aposta. Ao selecionarmos dentre o vasto universo entre as telas e as pessoas estes dois artefatos (*Memoryshare e Postsecret*) pensamos poder trazer alguns elementos para ampliar nosso conhecimento sobre as tendências em curso. Trata-se de propor aproximações que impliquem em redimensionar algumas questões teóricas e, ao mesmo tempo, propor aproximações entre objetos de caráter distintos (um mais comercial e outro um pouco mais artístico). Nestes projetos sobressai o caráter fragmentado e anônimo nos registros que remetem ao cotidiano dos indivíduos, registrados em uma plataforma social, e tornam o presente alvo de uma operação memorialista. Não se pode distinguir neles alguma temática comum a todos os registros, ao contrário do próximo caso que traz um outro modelo de plataforma digital que explora a circulação das memórias históricas na rede mundial de computadores.

2.3 O desgaste da documentação: revisitando o portal BBC People’s War³¹

Neste item apresentamos nossa reflexão sobre o arquivo digital People’s War da BBC. A análise deste caso é um desdobramento direto da nossa dissertação de mestrado, *O passado em bits* (Dantas:2008). Neste item não será analisado o projeto do, mas sim a possibilidade de explorar os processos de patrimonialização ali presentes através da documentação nascida digital. Por ter acompanhado a constituição do acervo e as modificações ao longo dos anos, consideramos oportuno uma análise a partir dos novos referenciais, mas sobretudo interessa-nos nos perguntar sobre o possível desgaste da documentação digital.

³⁰ Tradução livre: considerando as tendências múltiplas e contrárias e a tendência a uma pluralização do processo de rememoração nas sociedades, a conexão entre mídia e memória não pode ser descrita apenas de forma teórica, mas deve também ser observada empiricamente com minúcias.

³¹ Parte da discussão deste item foi publicada no artigo Passado e Presente nos registros digitais (DANTAS, DODEBEI:2009)

2.4 Os desgastes das memórias digitais

A seguinte passagem de William Gibson, um dos fundadores da literatura Cyberpunk³², do conto *Dead Man Speaks* (2203), nos ajuda a enveredar nesta reflexão sobre os pixels não encontrados:

“Time moves in one direction, memory in another.

We are that strange species that constructs artifacts intended to counter the natural flow of forgetting.

I sometimes think that nothing really is new; that the first pixels were particles of ochre clay, the bison rendered in just the resolution required. The bison still function perfectly, all these millennia later, and what screen in the world today shall we say that of in a decade? And yet the bison will be there for us, on whatever screens we have, carried out of the primal dark on some impulse we each have felt, as children, drawing. But carried nonetheless on this thing we have always been creating, this vast unlikely mechanism that carries memory in its interstices; this global, communal, prosthetic memory that we have been building since before we learned to build.” (GIBSON, 2003)³³

A partir desta provocação podemos nos perguntar quais seriam as telas onde os pixels aqui enfocados podem ser vistos em nossa prótese digital da memória? Ou dito de outra maneira: de que forma os projetos de memória, concretizado em práticas que transpassam as telas, podem ser acessados hoje? O MemoryShare e o PostSecret são plataformas que permanecem online, mas que sofreram algumas modificações ao longo do tempo. Por ser um projeto mais recente o MemoryShare não sofreu grandes mudanças visuais. De todo modo, averiguamos as possibilidades de pesquisa e verificamos que o website foi arquivado pelo UK Web Archive e pelo Internet Archive. Na página do arquivo britânico há uma versão inicial do projeto de 2009, enquanto que na WayBack Machine apenas pudemos encontrar uma documentação a partir de 2011, conforme mostram as linhas do tempo abaixo:

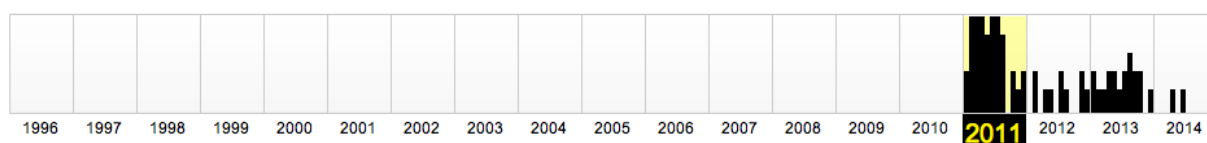
³² Sobre a importância da obra de William Gibson ver *Cybercultures Reader* (BELL, KENNEDY:2000).

³³ Conto publicado no Blog pessoal de William Gibson, disponível em: http://www.williamgibsonbooks.com/archive/2003_01_30_archive.asp (acesso em 12/05/2014)

Imagem 15 - Página do website *Internet Archive*



Fonte: UK Web Archive



Tela de linha do tempo das páginas contidas no Arquivo da Web das Ilhas Britânicas.

Fonte: Internet Archive

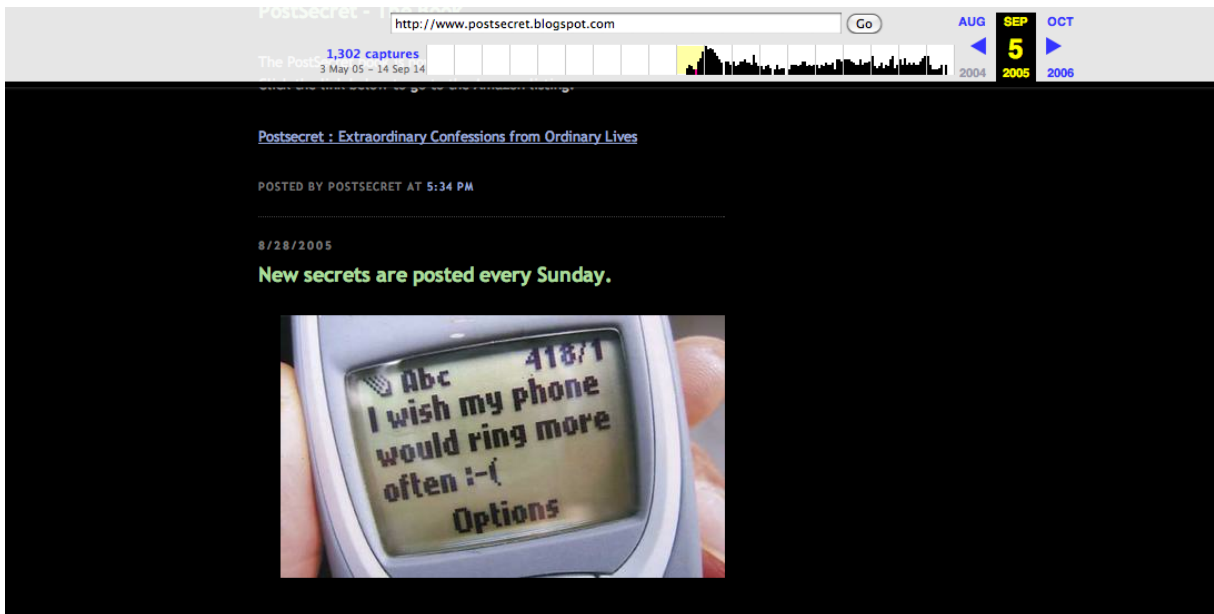
No caso do projeto PostSecret há um website paralelo com o objetivo de manter os conteúdos postados, mas não a interface. Ou seja, a forma de navegar é contemporânea, mas os postais foram preservados.

Imagem 16 - Arquivo do website *Post Secret*



Post Secret Archive (<http://postsecretarchive.com/2005/08/8-28-2005/>). Fonte: acervo pessoal da autora.

Imagem 17 - Registro do arquivo do website *Post Secret* no website *Internet Archive*



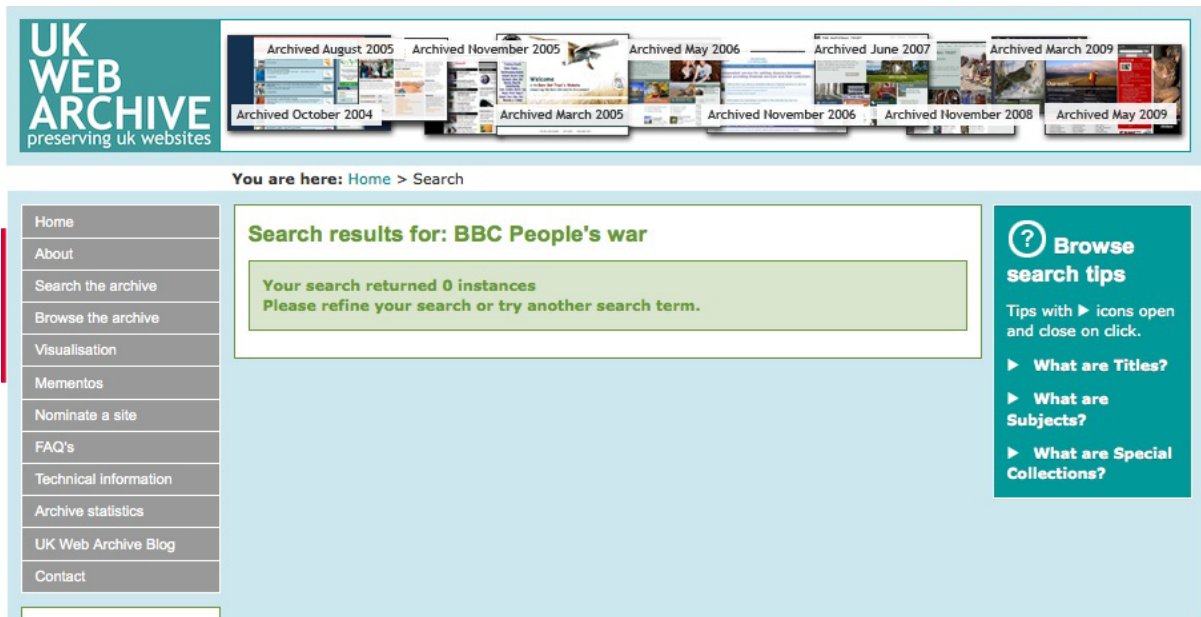
Tela do Postsecret arquivada pelo Internet Archive (<http://web.archive.org/web/20050905105408/http://www.postsecret.blogspot.com>). Fonte: acervo da autora.

Na imagem do PostSecret Archive, percebe-se que o postal é o mesmo, porém a interface de visualização é diferente. Ao acessar via Internet Archive pode-se ver o fundo preto, identidade visual do projeto, que não está mais presente no website criado. Por outro lado, no website mantido pelo projeto a interferência da publicidade é visível no banner acima da imagem (o texto está em holandês porque acessei a página dos Países Baixos) e também no box lateral. Este é um aspecto importante do projeto, que por outro lado, não está presente na página arquivada pelo Internet Archive (ou pelo menos, não estava acessível via WayBack Machine). A comparação entre essas duas imagens, ambas de agosto de 2005, revelam as distinções visíveis nas criptografias da memória.

No caso do projeto BBC People's War iniciamos nosso percurso de buscas no United Kingdom Web Archive, que se dedica desde 2004 a arquivar a web britânica³⁴. Na primeira busca pelo website não obtivemos resultado algum, como mostra a tela abaixo:

³⁴ Ver: <http://www.webarchive.org.uk/ukwa/> (acessado em 13/04/2014)

Imagem 18 - Resultado de busca no website *UK Web Archive*



Tela do resultado das busca pelo portal BBC People's War no UK Web Archive. Fonte Coleção da autora.

Porém, ao executarmos uma busca avançada, foi possível localizar uma captura da página do portal em 2007. O conteúdo em profundidade não pode ser acessado livremente online, mas imaginamos que uma consulta na British Library poderia implicar em um maior acesso aos dados, já que os sistemas entre arquivo e biblioteca ainda não estão articulados³⁵.

Imagem 19 - Registro do arquivo do website *WW2 People's War* no website *UK Web Archive*

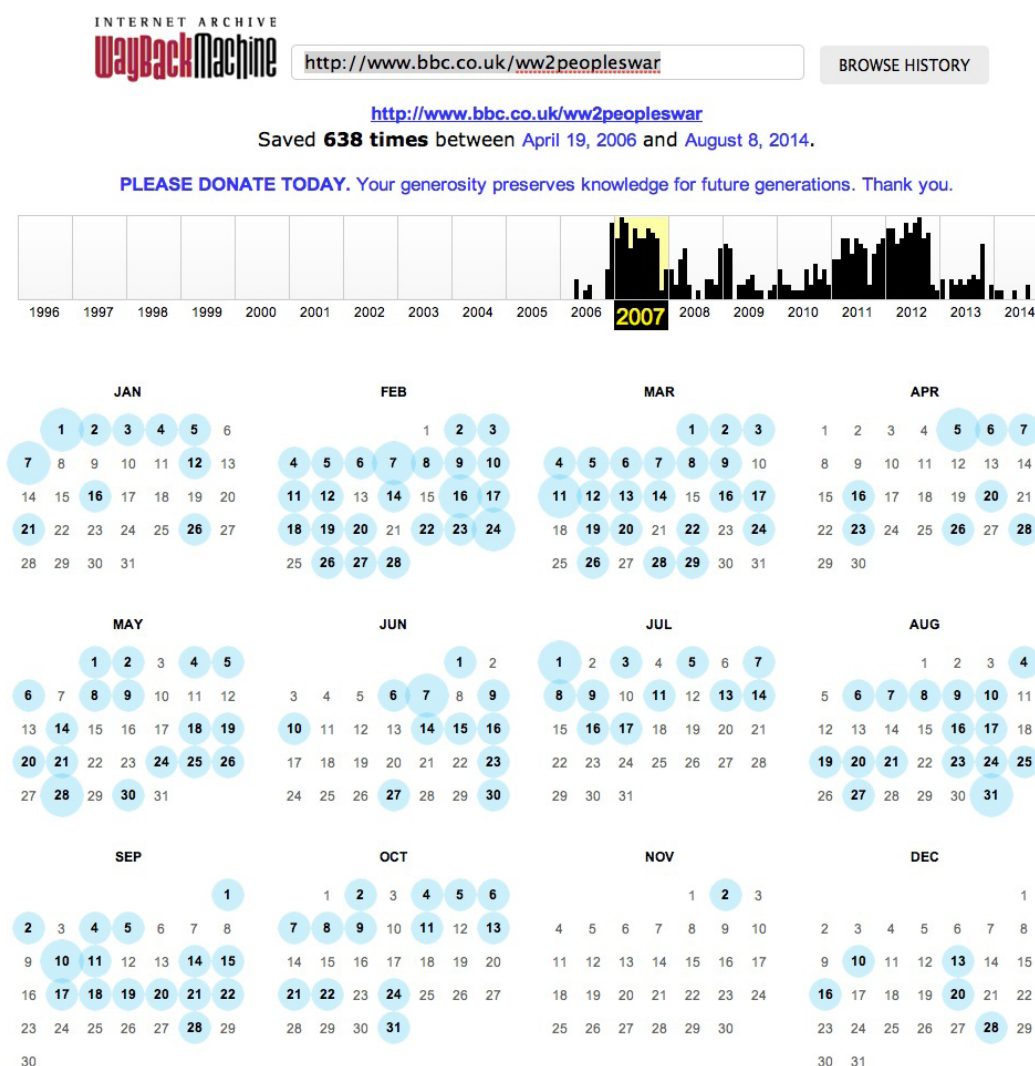


Tela de retorno UK Web Archive (<http://www.bl.uk/>). Fonte: coleção da autora.

³⁵ De acordo com a documentação do UK Web Archive há um esforço no sentido de articular os sistemas: “We work closely with colleagues across the Library on the ingest, storage and long-term preservation of web archives in the Digital Library System. Along with the Library’s other collections, discovery of archived websites is provided through the resource discovery system ‘Explore the British Library’, which is capable of finding records from both the Open UK Web Archive and the Legal Deposit UK Web Archive.” Disponível em: <http://www.bl.uk/aboutus/stratpolprog/digi/webarch/> (acesso em 10/05/2014)

A pesquisa revelou-se bem mais positiva ao realizarmos uma busca no Internet Archive, onde pudemos encontrar uma maior variedade de capturas do portal BBC People's War, graças ao fato de já possuímos o endereço eletrônico (URL) do projeto. Cada círculo azul claro no calendário abaixo indica uma data em que foi realizada a captura da página³⁶.

Imagem 20 - Resultado da busca pelo website *WW2 People's War* no website *Wayback Machine*



Tela contendo todas as capturas executadas pelo Internet Archive. Fonte: Internet Archive

³⁶ No próximo capítulo trataremos em mais detalhe sobre o funcionamento do Internet Archive, e, sobretudo, das formas de pesquisar na WayBack Machine.

Ao navegarmos nas ruínas do acervo digital da BBC arquivado pelo Internet Archive, percebemos que há um rico material a ser consultado. Essa documentação nascida digital traz lacunas igualmente importantes que lançam luz nas operações implicadas no armazenamento. A principal lacuna é a versão beta do site de 2004 e a versão do ano de lançamento 2005. Isto porque os formulários que guiavam as maneiras de participar, assim como os fóruns não foram preservados, ou pelo menos, não estão acessíveis online. Outra especificidade da documentação é ser de pouca profundidade, o que significa que os links contidos em cada captura não foram salvos³⁷.

Apontar essas ausências ou lacunas, próprias de qualquer conjunto documental, não significa adotar um discurso de que tudo deve ser salvo imediatamente. O alarmismo exacerbado nos parece contraproducente, porém é inegável que a poeira do arquivo digital pode ser até mais destrutiva do que aquela que se deposita vagarosamente sobre os papéis, ou sobre pedra, como provoca Gibson (2003). Este é o tema a ser desenvolvido no próximo capítulo, mas pensamos que indicar essas articulações poderia contribuir para uma visão mais panorâmica onde os projetos de memória mencionados tornam-se objeto do arquivamento sucessivo que podem ou não assegurar o acesso. As plataformas e interfaces que propiciam o acesso constituem-se igualmente em formas de esquecer ou lembrar, a questão é entender de que modo isto vem ocorrendo. Esta é a questão que será abordada no próximo capítulo.

37 Nas palavras de Janine Solberg: “the digital tools and structures that increasingly support our research efforts have material and epistemological implications for how we discover, access, and make sense of the past” (2012:1)

3. PÁGINAS E PROTOCOLOS ENREDADOS: A WEB E A CONSTITUIÇÃO DO SEU MAIOR ARQUIVO

A expansão das novas tecnologias da informação na última década do século XX foi paralela a uma discussão sobre as especificidades dessas mídias. O discurso sobre uma nova era das trevas⁴² engendrou uma série de iniciativas e discussões. Wolfgang Ernest (1999) considerou que a web seria um meio sem memória. Outros pesquisadores, porém, vêm trazendo importantes reflexões que chamam atenção para as dinâmicas entre o lembrar e esquecer, como o conceito de um “efêmero que perdura” (enduring ephemeral) de Wendy Hui Kyong Chun (2008:149) ou a versão de dupla face, acumuladora e amnésica (pak-rat or amnesiac) explorado por Belinda Barnet (2005). Pensamos que para compreender melhor a questão da memória digital, torna-se relevante elencar alguns momentos da história da web que foram constitutivos de sua arquitetura e, por conseguinte, das possibilidades de armazenamento e acesso aos dados.

Este capítulo está dividido em duas partes bastante distintas, mas com uma temática comum: o passado da web. Podemos pensar que as partes estão “enredadas” uma vez que trazem enredo (ou narrativa) e dados sobre a história da rede mundial de computadores. No primeiro item nosso desafio foi compilar algumas visões sobre a web, destacando alguns marcos cronológicos, mas sem arriscar uma narrativa ampla sobre a história da rede. Na verdade procuramos apenas elucidar aspectos relevantes para o entendimento de nossa problemática: a investigação sobre as criptografias da memória presentes nas práticas de arquivamento da web. Na segunda parte, tratamos do maior arquivo da web: o Internet Archive. Compreender como as formas de armazenamento e acesso foram desenvolvidas e, posteriormente, institucionalizadas é fundamental para qualquer empreitada no sentido de desvendar as criptografias da memória em meio digital.

3.1 Visões da rede: muito além do ciberespaço

Diante da impossibilidade de abordar toda a história da criação e desenvolvimento da rede mundial de computadores, optamos por focar a cronologia implicada na história da internet e da web, destacando, assim, aspectos relevantes da historiografia das tecnologias digitais. Trata-se, pois, de eleger alguns debates e não de apresentar os fatos sucessivamente, o que

⁴² Ver, por exemplo os textos e projetos de Danny Hillis, da Long Now Foundation. Disponível em: <http://archive.wired.com/wired/scenarios/clock.html>

considero uma opção válida e já muito bem realizada por obras fundamentais como, por exemplo, *A Galáxia da Internet*, do sociólogo de Manuel Castells (2003), onde é possível encontrar uma exploração panorâmica sobre as dimensões sociais e econômicas da internet.

A trilha escolhida neste trabalho é a de investigar alguns debates e chamar atenção para entrelaçamento das práticas e os artefatos tecnológicos. Trazer esta questão para o campo das criptografias da memória implica em perguntar-se sobre de que modo, por exemplo, a aclamada “memória dinâmica” se manifesta nas plataformas contemporâneas que permitem o acesso a variadas formas de documentação do passado. Ou, em polo oposto: inquirir de que maneira o discurso sobre as trevas digitais se faz presente. De fato estes são extremos que ainda pontuam algumas questões bastante atuais, seja nos processos de digitalização, seja naquilo que estamos mais interessados: o arquivamento da web.

Esta distinção entre um discurso corrente e as práticas relativas aos processos de circulação das memórias sociais pode ser bem elucidada em um trecho interessante do livro *Software Takes Command* (2008) de Lev Manovich. Ao tentar dar mais argumentos que validem a existência de um novo campo de saber, Software Studies, Manovich cita um trecho do livro *After the Great Divide*, de Andreas Huyssen (1986). Naquela passagem Huyssen⁴³ menciona como os bancos computadorizados da memória podem trazer de modo quase mágico imagens dos gêneros bastante distintos. Porém esta espécie de banco de imagens ainda não existia no momento em que Huyssen escrevia o trecho bastante profético. Manovich, que trabalhava na época em uma empresa de computação gráfica, afirma apenas mais de uma década depois, que algo semelhante poderia realmente existir quando programas de manipulação de imagens tornaram-se mais baratos e as estações de trabalho passaram a ter espaço de disco para suportar as operações. Interessante perceber que para Manovich havia apenas uma questão de deslocamento temporal na análise de Huysen, e que de fato a web é um banco de memória cultural acessível a todos, como se depreende da passagem abaixo:

“Only when the Web absorbed enough of the media archives it became this universal cultural memory bank accessible to all cultural producers. But even for the professionals, the ability to easily integrate multiple media sources within the same project – multiple layers of video, scanned still images, animation, graphics, and typography – only came towards the end of the 1990s.” (2008:96)

43 Eis a passagem mencionada de Andreas Huyssen: “All modern and avantgardist techniques, forms and images are now stored for instant recall in the computerized memory banks of our culture. But the same memory also stores all of pre-modernist art as well as the genres, codes, and image worlds of popular cultures and modern mass culture.” APUD, Manovich:2008:96

Na verdade é preciso contextualizar um pouco este trecho que diz respeito, sobretudo, ao uso das imagens por designers e artistas, remetendo a um quadro mais amplo de História da Arte. Ainda assim, consideramos problemático pensar na web como algo universal, uma vez que os conteúdos de acesso restrito (com pagamento) ainda são um impeditivo para o acesso livre às imagens. Além disso, com a ampliação das tecnologias de localização, caiu o mito do espaço desterritorializado. De todo modo o que esta discussão traz de mais interessante parece ser as distinções entre os discursos e as práticas envolvidas. Quando pensamos no caso do armazenamento das fontes digitais parece, algumas vezes, que estamos há anos luz deste banco integrado de herança cultural, uma vez que só em 1996 foram lançadas as principais iniciativas de preservação digital, e, como veremos mais adiante, a forma de recuperação da informação na época era de fato muito restrita.

A diversidade da produção acadêmica sobre o tema pode ser apreciada através das áreas de pesquisa em constituição como: Estudos de Software, Estudos de Internet, História das Novas Mídias, Estudos Sobre as Ferramentas de Busca, Arqueologia das Mídias, Humanidades Digitais... Cada área propõe uma certo viés de análise e pode-se notar um alargamento das discussões que vão se afastando das dicotomias simplistas entre os alarmistas e utópicos⁴⁴. Em um trabalho importante sobre a histórias da mídias, Briggs e Burke afirmam o seguinte:

não é possível nessa altura de sua história concluir que, pela facilidade de acesso e pela transformação a ‘partir de baixo’, ela desempenhará um papel renovador a longo prazo. Alguns críticos até temem que a internet mine todas as formas de autoridade, afete negativamente o comportamento e ameace a segurança individual e coletiva. Alguns especialistas em estudos de mídia, por conseguinte, tem posto em evidência de modo correto o que chamam de ‘debates de mídia’. Eles englobam temas específicos e processos a longo prazo que trazem novos dados e perspectivas(2006:4).

O entrelaçamento entre os debates em torno das novas mídias e as práticas a elas associadas podem revelar tensões e conflitos importantes. Por exemplo, durante mais de uma década o principal conceito em circulação para se pensar a rede mundial de computadores estava associado a metáfora do ciberespaço (STEVENSON:2013). De forma resumida podemos dizer que o termo se originou da obra *Neuromancer* (1982) de William Gibson. O autor integrava um grupo de escritores que criaram um gênero literário: o Cyberpunk. O termo, por sua vez, ao migrar das obras literárias para o discurso mais geral passou a receber diversas conotações,

⁴⁴ Para ser uma ideia a busca pela frase exata “software studies” retornou 2.712 itens de produção bibliográfica no Google Scholar. Pesquisa realizada em 10/04/2014.

nem sempre em consonância com o caráter distópico a ele associado em muitas das obras. Na verdade muitas vezes o conceito passou a carregar um conteúdo utópico, como na Declaração de Independência do Ciberespaço, publicada em 1996, por John Perry Barlow que se iniciava com extensamente citada afirmação:

“Governments of the Industrial World, you weary giants of flesh and steel, I come from Cyberspace, the new home of Mind. On behalf of the future, I ask you of the past to leave us alone. You are not welcome among us. You have no sovereignty where we gather”⁴⁵
(BARLOW:1996:1)

Um dos autores importantes para se pensar a extrapolação desta metáfora para a academia é o filósofo francês Pierre Levy. O conceito de tecnologias da inteligência, formulado por Levy, carrega um otimismo considerável em relação às implicações sociais das novas tecnologias. Não estamos sugerindo, no entanto, que os discursos são similares. A Declaração de Barlow tinha um caráter panfletário e circulou na web como uma forma de adesão às restrições comerciais e políticas de regulamentação do uso da rede. Já o pensamento de Pierre Levy, embora possua conexões com essa atmosfera que elevava o ciberespaço a quase um mundo paralelo, não pode ser reduzido a isto. Sua obra ainda constitui uma referência importante para que possamos compreender as interfaces digitais do nosso passado.

A perspectiva histórica adotada para analisar a informática é um dos pontos importantes da obra de Levy e advém de uma tradição de estudos de mídia que remontam a Harold Innis (1951) e Marshal McLuhan (1961), procurando, porém, atualizar as questões e discussões que foram por eles postuladas. É neste sentido que podemos compreender a preocupação de compor um panorama sobre as diferentes tecnologias da inteligência: oralidade, escrita e cibernética⁴⁶.

No entanto não podemos deixar de perceber o modo recorrente como é dado às mídias, ao tentar explicar seus aspectos sociais, uma espécie de autonomia frente aos grupos sociais. Explicando melhor, embora a reflexão de Levy (1999) tenha um caráter filosófico, a perspectiva histórica dali apreendida parece implicar em uma mola impulsionadora dos eventos e práticas inerentes às próprias tecnologias, uma vez que são narrados poucos conflitos ou experiências díspares. Esta tem sido também a tendência de uma gama de estudos.

45 Disponível em: <https://projects.eff.org/~barlow/Declaration-Final.html> (acesso em 10/04/2014)

46 É importante salientar que apesar do otimismo significativo, Pierre Levy aponta também para a continuidade e em várias passagens reafirma a ideia de que “*as técnicas não determinam, elas condicionam. Abrem um largo leque de novas possibilidades das quais somente um pequeno número é selecionado ou percebido pelos atores sociais*” (1999:20)

Em um balanço historiográfico sobre a historiografia das mídias, Lisa Gitelman, em *Always Already New: Media, History and Data of Culture* (2008), aponta para a existência de uma visão bastante generalizada de compor as narrativas sobre a história das mídias a partir de um fluxo de consequências ou efeitos. O problema, aponta a autora, não é simplesmente perguntar-se sobre quais os efeitos sociais engendrados pelo uso de determinadas tecnologias, mas sim de, reversamente, explicar as práticas sociais tão somente pelo fator tecnológico (2008:61).

A narrativa da invenção do gramofone, muitas vezes centrada na evolução técnica, que não levava em consideração o inicial fracasso na comercialização da invenção de Thomas Edison, é analisada pela autora. A primeira utilização proposta para o uso do gramofone seria como um equipamento de escritório que possibilitasse o ditado, e depois, a cópia pelas secretárias. Porém esta proposta não foi bem sucedida e, apenas quando os primeiros usuários começaram a subverter essa proposta, utilizando para gravar músicas, foi que o aparelho conquistou o mercado. A elucidação deste processo histórico, demonstra a necessidade de se romper com um eixo unicamente evolutivo, mas não desconsidera a existência de uma lógica intrínseca ou uma materialidade das mídias. É um jogo delicado e pendular, entre o negar o determinismo tecnológico, sem descuidar de investigar os condicionamentos. Nas palavras de Gitelman:

“Media are more properly the results of social and economic forces, so that any technological logic they possess is only apparently intrinsic. That said, I have also resisted taking a reductively antideterministic position. At certain levels media are very influential, and their materiality properties (literally and figuratively) matter, determining some of the local conditions of communication amid the broader circulations that at once express and constitute social relations” (2008:10)

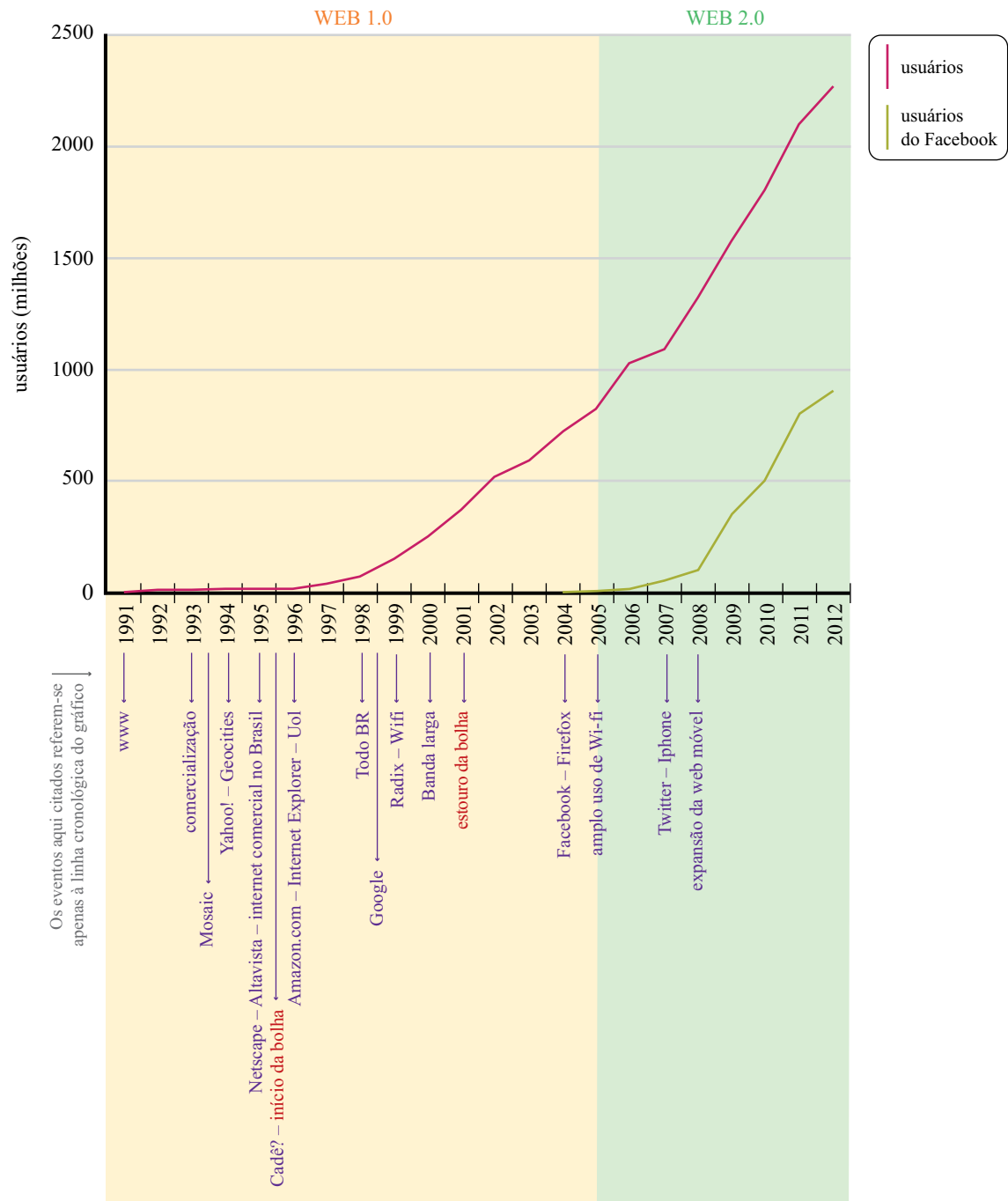
Se pensarmos nas práticas contemporâneas de arquivamento da web, a partir desta perspectiva, podemos perceber que embora o desenvolvimento tecnológico permita uma série de operações e usos, há significativas restrições sociais. Por exemplo, a utilização de programas sofisticados de arquivamento prevê uma expansão dos treinamentos dos profissionais envolvidos nas instituições de preservação da memória, o que não ocorre de modo semelhante em todos os países. Por outro lado, os fatores econômicos associados a manutenção de data centers muitas vezes impossibilitam a execução de determinados projetos. No caso do Brasil, que tratamos no primeiro capítulo, defendemos que, mais do que uma questão meramente tecnológica, há que se dar destaque aos conceitos de patrimônio digital em circulação que privilegiam a digitalização e não acervos nascidos digitais. Esta é apenas uma derivação do debate presente na pesquisa de Lisa Gitelman que traz reflexões interessantes para o nosso trabalho. Mas, cabe ampliar um pouco o foco no entendimento das narrativas sobre a história da rede.

3.1.1 A retórica da ruptura na cronologia da web

Como entender a superposição entre uma utopia democratizante com um determinado meio de comunicação? Tendo em perspectiva a recente história de alguns outros meios, que antes foram igualmente novas mídias, como o cinema e a televisão, pode notar que não houve uma associação com um imaginário libertário. Fred Turner, em *From Counterculture to Cyberculture* (2006), explorou minuciosamente a mencionada superposição, elaborando uma crítica perspicaz à mitologia relacionada ao ciberespaço. De acordo com sua pesquisa o ideário contracultural acoplou-se às metáforas relativas à rede mundial de computadores por vários fatores, entre eles o comercial, uma vez que o discurso de que a web seria benéfica às liberdades individuais não deixou de repercutir na venda do aparelho para lares e escritórios de todo o mundo.

A proposta de pensar historicamente sobre web implica em apontar de que modo tais utopias foram incorporadas pelas empresas. Michael Steveson (2013) denomina este imaginário de “excepcionalismo da web” uma vez que pretenderia afirmar que o próprio traria uma ruptura nos modelos de comunicação. As publicações na web como Wired, Mondo 2000 e Slashdot, divulgavam um ideário simplista que procuravam definir a criação da web como evento mais revolucionário da história. Esta retórica ficou conhecida como cyberbole e não deixaria de angariar muitos críticos. Mais recentemente Eugeny Morozov abordou, em *Net Delusion* (2011) o que ele classificou da grande desilusão causada pela rede, que estaria justamente associada a não concretização deste imaginário. Se o termo ciberespaço representou o discurso utópico na década de 1990, a partir de 2004/2005 teria início um outro ciclo de metáforas. O quadro abaixo procura dar uma dimensão destes períodos:

Imagem 21 - Gráfico relacionando usuários das web 1.0 e 2.0 à quantidade de usuários do Facebook



Fonte: Adaptação de Schuler (2012) com acréscimo de informações de Carvalho (2006), RNP (1998).

Há um relativo consenso quanto a esta cronologia, embora os debates quanto as definições e qualidades dos termos ainda não estejam nem um pouco definidos (SCHULER, 2012). A web 1.0 seriam os primeiro anos da Internet até 2005, quando as plataformas de inserção de conteúdo colaborativo se expandiram, como por exemplo o youtube e o facebook. Aliás, uma

das importantes características da web 2.0 seria justamente a ampliação de um circuito de comunicação de reciprocidade entre todos os pontos da rede. Mas, podemos nos perguntar: qual teria sido a grande ruptura? De acordo com Stevenson (2012), na verdade algumas práticas de construção colaborativa são a base da arquitetura da web, tal qual pensada por Tim-Berners-Lee, o que houve foi uma massificação destes procedimentos com fins mercadológicos.

Se desde 1982, com o estabelecimento do protocolo de transferência (TCP-IP) a internet já estava em plena atividade, a web (WWW) só iria se concretizar a partir de 1990. No início da década de 1980, o interesse comercial na rede já se fazia presente através da atuação de provedores de acesso que faziam a conexão de pessoas físicas com a rede nos EUA. A ampliação da rede fizeram surgir uma série de dispositivos novos, mas não resta dúvidas que o principal deles foi mesmo uma criação vinda da Europa, no CERN, um laboratório suíço de pesquisa de física de partículas onde o inglês Tim Berners-Lee criou as bases para o desenvolvimento da World Wide Web.

Mas, afinal, qual é a diferença entre a internet e a web? Seguindo as informações contidas nos documentos oficiais do World Wide Web Consortium (W3C) podemos dizer que a internet é a rede das redes, construída a partir do protocolo TPC/IP enquanto a web é um espaço informacional onde a troca de informações ocorre via URI (Uniform Resource Identifiers). Eis o texto do W3C:

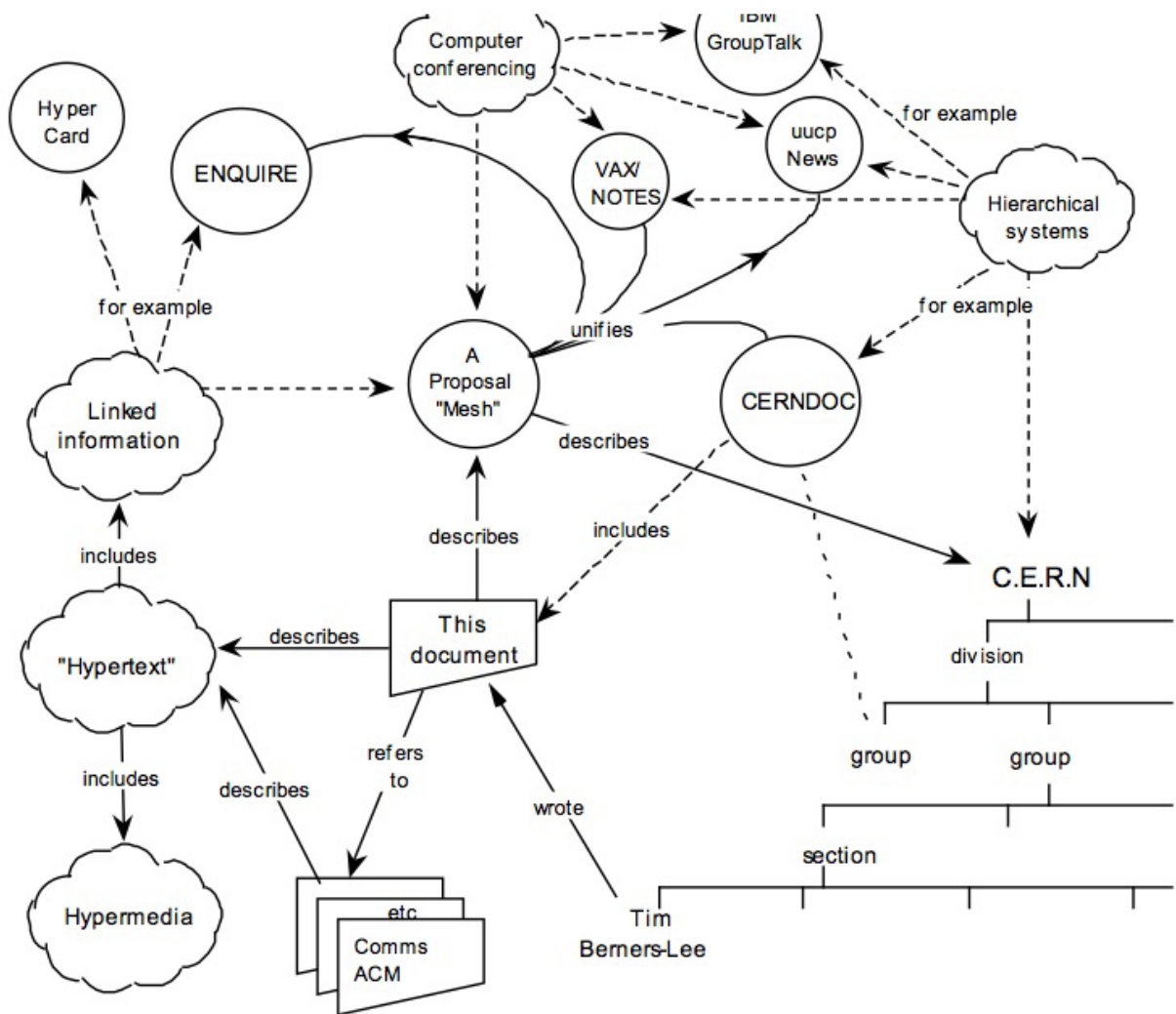
“The World Wide Web (WWW, or simply Web) is an information space in which the items of interest, referred to as resources, are identified by global identifiers called Uniform Resource Identifiers (URI).(..) A defining characteristic of the Web is that it allows embedded references to other resources via URIs. The simplicity of creating hypertext links using absolute URIs () and relative URI references (and) is partly (perhaps largely) responsible for the success of the hypertext Web as we know it today.” (W3C:2004)⁴⁷

Deste modo, a troca mútua e irrestrita de conteúdo entre computadores sem pertencimento a uma rede fechada, bem como a criação de protocolos e linguagens comuns foram a base para a rede mundial de computadores, ou simplesmente, a web. A possibilidade de clicar em *hiperlinks* nas páginas dos documentos e saltar para outras páginas foi um dos procedimentos centrais pensados por Berners-Lee e a equipe do CERN e implementados com a estruturação de uma linguagem específica para web a HTML (*Hypertext Markup Language*). Além disso, a criação de uma forma de identificar as máquinas através de um protocolo, primeiramente

⁴⁷ Disponível em: <http://www.w3.org/TR/webarch/> (acesso em 10/04/2014).

o URL (*Uniform Resource Locator*) e depois o HTTP (*Hypertext Transfer Protocol*) foram desenvolvimentos que tornaram a Web possível⁴⁸. A imagem abaixo integra o artigo “Information Management: a proposal” escrito por Tim Berns-Lee em 1989 e como se pode ver é bem mais que uma ilustração. É na verdade uma narrativa visual sobre o próprio documento e as propostas nele contidas. Não conclui, mas afirma entre as setas e balões as continuidades entre a chamada web 1.0 e a web 2.0.

Imagem 22 - Esquema afirmando as continuidades entre web 1.0 e web 2.0



Fonte: BERNERS-LEE (1989:1)

⁴⁸ Ver, documento do The World Wide Web Consortium (W3C), consórcio criado em 1994 com o objetivo de padronizar a emergente World Wide Web (<http://www.w3.org/WWW/> : acessado em 05/04/2014).

Entre a proposta de Tim Berners-Lee e a efetiva utilização do modelo passaram-se alguns meses. Mas, a incorporação por um público maior, fora dos ambientes acadêmicos e militares, ainda demoraria alguns anos. A associação entre o barateamento dos computadores para uso pessoal e o desenvolvimento de alguns recursos informacionais, como o navegador Netscape, em 1995, possibilitou a utilização da rede por pessoas não familiarizadas com as linguagens da informática. A supremacia da Netscape permaneceu até o avanço do Internet Explorer em 1998. As operações monopolistas da Microsoft levaram a disputas judiciais⁴⁹. Assim, dos primórdios do Mosaic (a primeira interface de navegação), depois Netscape, chegaríamos ao reino da supremacia questionável, mas inegável, do Internet Explorer da Microsoft (CASTELS, 2001:45).

3.1.2 A ampliação da rede e os oráculos digitais

Para acessar a web, além do navegador, tornou-se fundamental o papel das ferramentas de busca. Como achar a informação necessária? Hoje, com um celular conectado à internet podemos consultar uma ferramenta de busca e acessar em segundos a previsão do tempo uma enciclopédia online. Porém na década de 1990, este não era bem o cenário. Embora a metáfora do oráculo já fosse utilizada para algumas ferramentas de busca, é preciso sinalizar que até o final daquela década as principais ferramentas ainda utilizavam índices organizados manualmente. Se hoje, nem tudo está acessível, mas é preciso reconhecer a novidade de ter na ponta dos dedos o acesso a uma enorme gama de informações.

Um das primeiras experiências de ferramenta de busca foi o Archie, um programa criado por Alan Emtage, que procurava organizar os principais diretórios já disponíveis na web. As principais ferramentas de busca do final da década de 1990 incluíam o Yahoo, Altavista e Lycos (AMARAL, 2009:3). Eram chamados sistemas de diretórios, indexados manualmente que utilizavam uma organização hierárquica das categorias.

As ferramentas de busca tornaram-se um filtro importante para os conteúdos da rede mundial de computadores. Em 1996, os estudantes de pós-graduação da Universidade de Stanford, Larry Page e Sergey Brin iniciaram um projeto de pesquisa sobre a relação matemática de links na internet (VAIDHYANATHAN, 2012:5). Uma versão de teste foi colocada online pela

⁴⁹ Sobre a história da Netscape, ver: Castañer, Xavier. *Competing on Internet Time: Lessons from Netscape and Its Battle with Microsoft*. Universidade de Minnesota, 2000. Disponível em: <http://webpages.csom.umn.edu/smo/smrc/REV32.pdf>. (acesso em 10/04/2014)

primeira vez em 1997 e obteve uma rápida repercussão no meio acadêmico e depois entre os usuários da internet como um todo. O sistema proposto pelo Google rompia com o sistema de organização hierárquica e buscava recuperar e organizar os links a partir de uma lógica da própria web. De acordo com Geert Lovink:

“Nowadays an altogether new phenomenon is causing alarm: search engines rank according to popularity, not truth. Search is the way we now live. With the dramatic increase of accessed information, we have become hooked on retrieval tools. We look for telephone numbers, addresses, opening times, a person’s name, flight details, best deals and in a frantic mood declare the ever growing pile of grey matter ‘data trash’. Soon we will search and only get lost.” (Lovink, 2012:1)

Em 2006, o verbo “to google” foi adicionado ao dicionário Webster , e passou adá língua inglesa, mas expandiu-se pelo mundo e tornou-se sinônimo de busca de informação. Neste ponto é preciso um esforço de incorporação dos referenciais teóricos propostos ao pensarmos sobre a escalada desta empresa. O sucesso do Google foi tão estrondoso que há uma tendência em focar exclusivamente na invenção de Page e Brin. A fascinante história de um algoritmo criado pelos estudantes de Stanford que permitia uma organização dos resultados a partir do número de *links* externos a um certo site, muitas vezes é narrada como a caça ao ouro do milênio. Hoje na maioria dos países ocidentais o Google é a ferramenta de busca mais utilizada, mas a escalada monopolista deve ser compreendida a partir da emergência de uma indústria das buscas. Em *Search Engine Freedom* traça um histórico deste processo como fica elucidado na citação abaixo:

“The eventual consolidation of the search engine market can be historically tracked to a number of contributing factors. Some of those factors are inherent in the operation of a general purpose search engine, such as the growing, evolved expectations of search services by end-users or the grown barriers to entry into the market. Other contributing factors include the integration of the search engine as an important business asset in the digital media and ICT industry, thereby reproducing existing consolidation in related markets in the search engine context. Of the many search companies that started offering their service in the 1990s, only Google remains as a mature independent company with its own search engine at its heart. (HOBOKEN, 2009:32)

Como fica explícito na passage acima, trata-se da instituição de um Mercado que passaria a exercer uma enorme influência em vários aspectos sociais. Não seria possível explorar de modo mais aprofundado a trajetória do Google dentro do escopo deste trabalho, porém a sua importância deve ser mencionada. O principal aspecto relacionado ao nosso trabalho é justamente a política de empresa de preservar todos os dados das buscas em seus servidores. Com a ampliação da capacidade de armazenamento da empresa, isso significou uma escalada na retenção de dados que vem sendo amplamente criticados. A expansão da coleta de dados implicaria em uma negação da possibilidade do esquecimento produtivo. Segundo, por exemplo, Victor Mayer-Schonbergh (2009). Porém o caráter comercial do empreendimento, que se reflete nos algoritmos e nos rankings dos resultados, implicam em um tipo de seletividade própria desta indústria. Nesta pesquisa não iremos enfrentar as criptografias envolvidas na googlização da memória cultural, embora devemos reconhecer que é um tema que deve ser cotejado.

3.2 Enredos e dados do Internet Archive

A proposta deste item é a de analisar o Internet Archive, o maior arquivo de da web, que em agosto de 2004 possuía nada menos que mais de 400 bilhões de páginas arquivadas. Mas como entender este arquivo? Ou melhor: seria o Internet Archive um arquivo? Para tentar nos aproximar destas questões iremos, primeiramente, tratar do projeto da instituição relacionando-o ao ideário do ciberespaço e da idade das trevas digitais.

O maior arquivo de páginas da internet foi criado, em abril de 1996, pelo engenheiro da computação Brewster Kahle a partir da base de dados da Alexa, sua empresa de recuperação de informação da web. As bases de dados com a varredura dos dados da web começariam a compor o acervo apenas no final daquele ano (KAHLE, 1998).

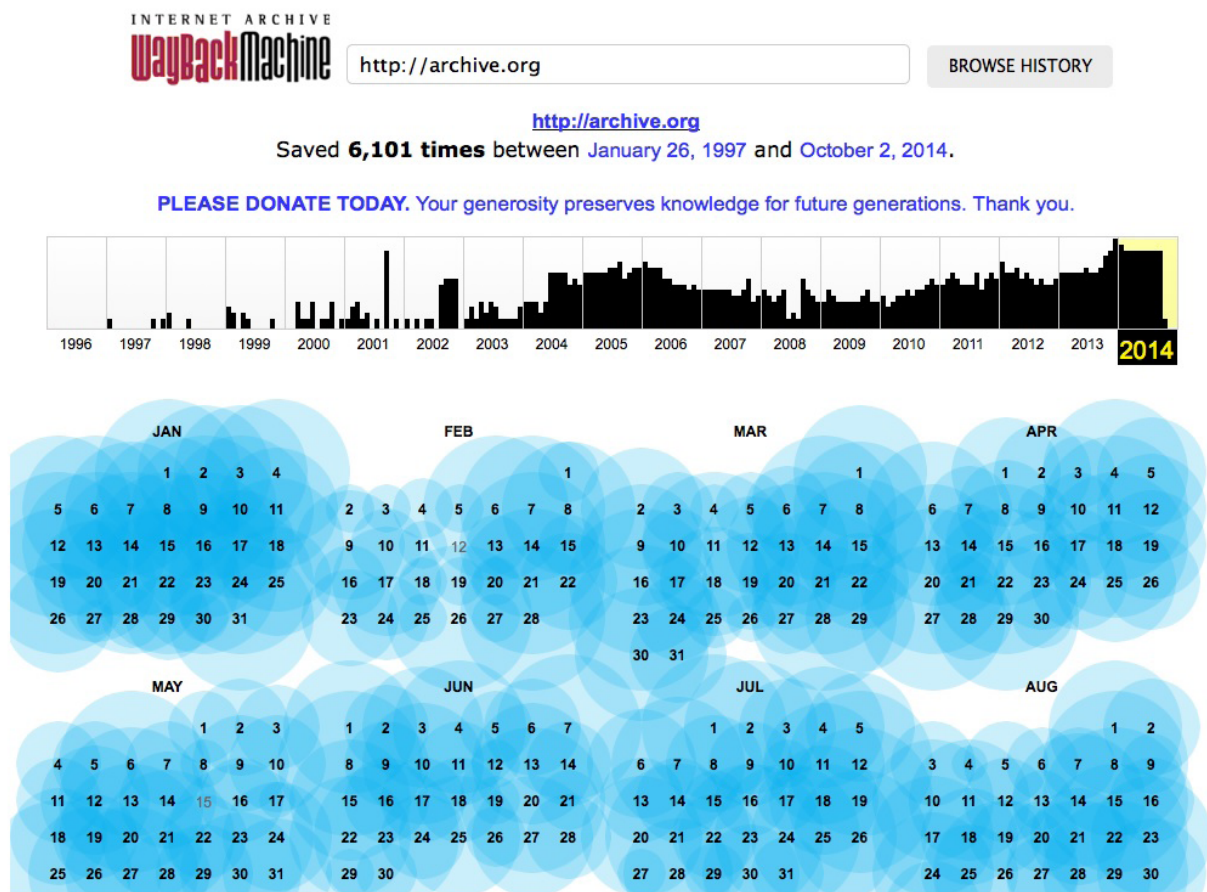
O Internet Archive é uma fundação sem fins lucrativos, associada a vários órgãos de preservação como a UNESCO e a Biblioteca do Congresso dos EUA. A cada duas semanas são adicionadas aproximadamente um milhão de páginas ao acervo do Internet Archive⁵⁰. Além de arquivar a própria web, o arquivo se estabeleceu como uma plataforma de preservação da mais variadas mídias: livros, imagens em movimentos, softwares, músicas, etc. Nosso foco neste capítulo está no arquivamento de páginas da rede mundial de computadores.

⁵⁰ Khale, Brewster. Entrevista ao programa Triangulation. 10/04/2013. Disponível em: <https://www.youtube.com/watch?v=06CkHpuGLEM>.

3.2.1 O projeto do Internet Archive: legado cultural em dados

O primeiro registro online do Internet Archive, arquivado pelo próprio, é de 26 janeiro de 1997, onde a única possibilidade de consulta ao acervo seria através de uma coleção sobre as eleições nos EUA, com curadoria do Museu Smithsonian⁵¹. A imagem abaixo nos indica que o próprio arquivo executou poucas capturas da sua página pública até o ano de 2002, quando paulatinamente verifica-se a inclusão de mais capturas.

Imagem 23 - Resultado da busca pelo site *archive.org* nos registros do website *Wayback Machine*



Fonte: Internet Archive.

⁵¹ Disponível em: <http://web.archive.org/web/19970126045828/http://www.archive.org/> (consulta em 10/04/2013)

A análise sobre o Internet Archive que se segue tem como objetivo elucidar o paradoxo entre a volatilidade dos registros na web e a possibilidade de armazenamento que o meio oferece. Para tanto seguiremos um percurso de descrição da plataforma, partindo das interfaces do arquivo e procurando dar a ver as escolhas técnicas.

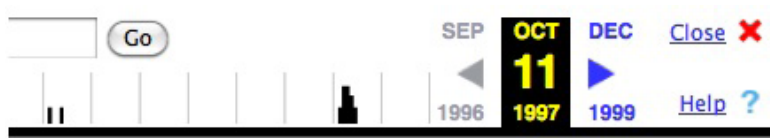
Logo de início, nas páginas de apresentação do Internet Archive, lê-se o seguinte:

“At present, the size of our Web collection is such that using it requires programming skills. However, we are hopeful about the development of tools and methods that will give the general public easy and meaningful access to our collective history. In addition to developing our own collections, we are working to promote the formation of other Internet libraries in the United States and elsewhere.”⁵²

A primeira página arquivada no Internet Archive da própria instituição data de 26 de janeiro de 1997. Navegando nesta página arquivada percebe-se a escala do projeto naquele momento. Não mais de 10 links integram a página inicial. Não há nenhuma possibilidade de acesso direto ao conteúdo arquivado. O projeto é apresentado através dos textos de Brewster Kahle.

52 Disponível em: <http://web.archive.org/web/20110514001454/http://www.archive.org/about/about.php> (acessado em 15 de julho de 2015)

Imagem 24 - Registro de 1997 do website *Internet Archive* no próprio website



Building TM INTERNET ARCHIVE a digital library for the future



[Acknowledgements](#)
[Board Members](#)
[Finding Us](#)
[In the News](#)
[Webmasters](#)

Our Mission

Internet Archive is collecting and storing public materials from the Internet such as the World Wide Web, Netnews, and downloadable software which have been donated by [Alexa Internet](#).

The Archive will provide historians, researchers, scholars, and others access to this vast collection of data (reaching ten terabytes), and ensure the longevity of this information.

For more information about our philosophy and objectives, please read [Archiving the Net](#) by the Archive's founder, [Brewster Kahle](#).

Visit: [The '96 US Presidential Election Web Archive](#). This archive was created in affiliation with the [Smithsonian Institution](#).

✉ info@archive.org

Fonte: Internet Archive

Cabe nos perguntar sobre qual seria então a proposta do Internet Archive? Qual foi a estratégia proposta para ir de encontro a mencionada natureza fugidia da informação contida na web? Para responder a esta pergunta inicial nos propomos a uma leitura atenta dos documentos fundadores do Internet Archive feita em paralelo com uma análise dos conteúdos do portal online neste período. Consideramos como documentos fundamentais os seguintes artigos: *Archiving the Net*⁵³ (Kahle,1997) e *Archiving Digital Cultural Artifacts: organizing*

53 Scientific American, Issue March 1997. Disponível em: http://web.archive.org/web/19971011050140/http://www.archive.org/sciam_article.html (acesso em 10/02/2014)

*an agenda for action*⁵⁴ (LYMAN, KAHLE, 1998). Estes dois artigos possuem uma estrutura narrativa similar no que se refere a defesa de um projeto de resgate de uma herança cultural em risco. O elemento central da narrativa encontra-se na eminente perda de todo o conhecimento gerado em suporte digital.

O principal argumento apresentado como justificativa para necessidade da criação de um arquivo da rede mundial de computadores partia da premissa de que a velocidade de substituição das páginas levaria a perda dos conteúdos publicados em suporte digital. Na época a taxa de substituição das páginas da web era calculada em torno de 75 dias (KAHLE e LYMAN, 1998:12).

O nome da empresa fundada por Kahle, Alexa, já antecipa o ideário a ser acionado. A Biblioteca de Alexandria é um símbolo fundamental nas narrativas para construção de um patrimônio bibliográfico no ocidente (WARWICK, 2014) e foi também o ponto de partida escolhido por Kahle para construir a seu argumento em defesa da preservação dos artefatos digitais. O texto inicia-se com a constatação da perda de parte dos manuscritos de Alexandria inserindo-a num contínuo onde outros bens culturais desapareceram, como os primeiros impressos e as películas cinematográficas do inícios do século XX. Neste ponto Kahle prossegue em tom drástico ao afirmar que milhões de páginas estão sendo perdidas e que não há ninguém preservando as páginas da rede mundial de computadores. O Internet Archive surge portanto para solucionar este problema e garantir a existência de uma “biblioteca digital” universal.

Em *Archiving Digital Cultural Artifacts: organizing an agenda for action*, Kahle e Lyman (1998) afirmam de maneira categórica a necessidade de ação ante a fugacidade do suporte digital, em especial da web:

“Digital information is seemingly ubiquitous as a medium for communication and expression, increasingly strategic for scientific discovery and the records that constitute institutional order in a modern society. And yet, it is at the same time fugitive, the pace of technical change makes digital information disappear before we realize the importance of preserving it. Like oral culture, digital information has been allowed to become a medium for the present, neither a record of the past nor a message to the future. Unless, that is, we redesign it now.” (1998: 5)⁵⁵

Nesta argumentação em defesa da preservação dos conteúdos digitais traz um diagnóstico pouco otimista sobre as possibilidades técnicas para tal empreendimento. Em 1997, a empresa

⁵⁴ Dig—Lib Julho/Agosto 1998. Disponível em: <http://www.dlib.org/dlib/july98/07lyman.html>






⁵⁵ Idem.

já contava com um milhão e meio de páginas perdidas, ou seja, endereços (URLs) que não mais existiam online. Como armazenar e, sobretudo, acessar tal conteúdo? Inicialmente a Alexa repassava ao Internet Archive uma base de dados de varredura da web com um intervalo de tempo de seis meses, e a instituição garantia a preservação, mas não o acesso.

Imagem 25 - Registro de 2000 do website *Internet Archive* no próprio website

The screenshot shows the Internet Archive website interface from 2000. At the top, there is a calendar navigation for June 2000. The main header reads "The Internet Archive: Building an 'Internet Library'". Below the header are navigation links: "Home", "News", "Jobs", and "Contacts". A secondary navigation bar includes "In the Collections", "Using the Collections", and "About the Archive".

The main content area features a section titled "How Big Is 14 Terabytes?" with the text: "Here's how the size of the Archive's collections today — containing material dating from 1996 to the present — compares to some familiar data banks:". Below this is a table with 10 rows, each containing an image, a description of a data bank, and its size in terabytes or gigabytes.

	A video store (5,000 videos)	8 terabytes (1 gigabyte per hour of video)
	A radio station (10,000 LPs and CDs, or 15,000 hours of music)	8 terabytes (535 megabytes per hour of music)
	A copy of your favorite mystery novel	1 megabyte
	One copy of the the <i>Encyclopaedia Britannica</i> (2,619 pages per copy)	1 gigabyte
	A thousand copies of the <i>Encyclopaedia Britannica</i>	1 terabyte
	A public library branch (300,000 books)	3 terabytes
	The ancient Library of Alexandria (400,000 scrolls)	800 gigabytes
	The Internet Archive's Web collection in March 2000 (about a billion Web pages)	13+ terabytes (and growing at a rate of 10 percent a month)
	The Library of Congress (20 million books, not counting pictures)	20 terabytes

Fonte: Internet Archive

Cabe, porém, a pergunta: mas armazenar tudo seria similar a não processar a dinâmica do esquecimento? O projeto de Brwster Kahle recebeu logo no início uma série de críticas na imprensa, e, mais recentemente, alguns pesquisadores também apontaram a estranheza de tal projeto. A associação entre o Internet Archive e Funes pode ser lida, justamente, como uma

crítica ao empreendimento. Já em julho de 1996, em um artigo da popular revista Slate, *The art of loosing*, o jornalista David Berreby faz um ácida crítica ao projeto, então apenas iniciando: “The thought of capturing all this useful information for future scholars is appealing, at least at first glance. (...) But with Borges in mind, I can’t really wish Kahle good luck and baud speed”⁵⁶. Um dos principais argumentos apresentados no texto em crítica ao projeto é o de que os objetos a serem preservados não eram valiosos: páginas de adolescentes, banalidades, registros mundanos de um cotidiano online seriam preservados com qual objetivo? Percebe-se que o autor possui uma visão de legado cultural alicerçado em uma tradição de seletividade rigorosa, onde várias instâncias sociais escolhiam aquilo que deveria ser preservado. O único ponto que sobressai ao ler o artigo com os olhos de 2014 é a pressuposta crença na viabilidade do projeto. Ou seja, o jornalista assume a proposta Internet Archive como algo viável. Hoje porém fica evidente a não totalidade deste acervo. Nem todas as páginas foram salvas, e as que estão incluídas estão longe de estarem completas. Funes não se personificou nos servidores do Internet Archive.

A metáfora da biblioteca digital como o elo possível entre o passado e o futuro passa a exercer uma centralidade nos discursos sobre a memória no final do século XX. Manoff (2010) descreve o cenário no qual o Internet Archive foi pensado nos seguintes termos:

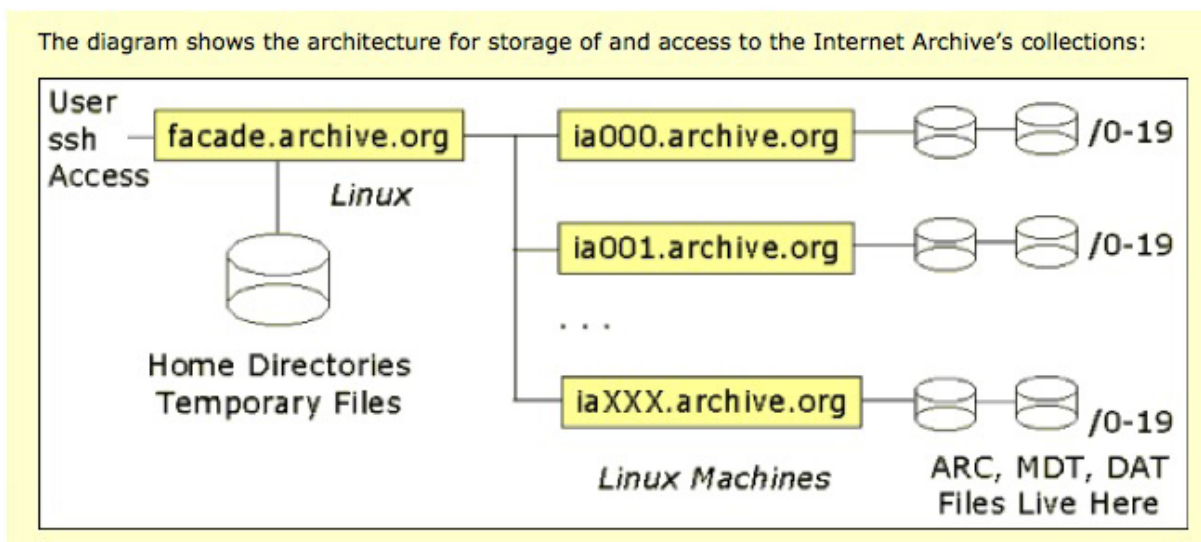
“In the cultural imaginary, the library stands as a bulwark of history, culture, and memory. (...) On the one hand, computers are seen as the ultimate memory tool capable of providing access to everything ever written. On the other hand, the digital record is liable to manipulation, distortion or erasure. In such an environment, we cling to our libraries and our artifacts even as we dream of ever more encompassing virtual collections.” (2010:379)

3.2.2 *A máquina do tempo do Internet Archive*

O software de busca, denominado Way Back Machine (WBM), foi criado em 2001 o que tornou viável a consulta ao acervo por meio de endereços (URL). Antes disso a consulta era realizada primeiramente entrando-se em contato com a instituição a fim de estabelecer uma forma de acesso. Porém até a criação do software WBM a possibilidade de consulta exigia habilidades de programação em UNIX. O diagrama abaixo dá uma dimensão da arquitetura informacional do Internet Archive.

⁵⁶ Disponível em: <http://www.unz.org/Pub/Slate-1996jul-00032> (acesso em 10/04/2014)

Imagem 26 - Diagrama: a arquitetura para registro e acesso às coleções do website *Internet Archive*



Fonte: Internet Archive

O esquema demonstra a preocupação com a duplicidade dos dados com fins de manutenção do acervo. O que chama atenção são os tipos de formato escolhidos: ARC, MDT e DAT. Embora não seja possível uma incursão do ponto de vista da tecnologia da informação, é importante perceber aí o surgimento de uma arquitetura voltada para a preservação do objeto nascido digital. O formato ARC foi criado pelo Internet Archive para viabilizar o empacotamento de uma massa de dados ao mesmo tempo que garantiria o armazenamento de metadados, como data da captura e endereço do servidor acessado (KAHLE:1996)⁵⁷.

Em 2001 foi criado um software, chamado Wayback Machine, desenvolvido pela Alexa para recuperar informações de um index de 10 bilhões de documentos (100TB). A arquitetura deste software que não é totalmente aberta e demonstrou ao longo do tempo algumas especificidades problemáticas. Do ponto de vista computacional há pouca flexibilidade das linguagens de programação utilizadas, o que de acordo com Brad Tofel (2007;5) traz dificuldades para inclusão de novas funções. Algumas especificidades da arquitetura do software, relacionadas ao modo de captura e retorno das buscas, devem ser mencionadas como forma de melhor compreender como as páginas são processadas e acessadas. Um dos aspectos que mais me chamou atenção ao pesquisar os documentos via WayBack Machine eram as camadas de temporalidades e os saltos no tempo que pareciam pipocar em muitos documentos. Assim, embora eu achasse estar no comando de uma máquina do tempo, e definisse um marco cronológico exato, ao abrir

⁵⁷ A documentação sobre o formato ARC está disponível em:

as páginas de 1998, muitas vezes eu via surgirem pop-ups muito mais recentes, até mesmo do dia e da hora em que eu estava espiando os tais dígitos. Entender este truque, ou este aspecto específico, deste documento digital foi um impulso inicial na direção de artigos técnicos. Este é um dos aspectos em que a noção de criptografias da memória parece se adequar plenamente, uma vez que é necessária uma certa decodificação para acessar os dados.

A seguir apresentarei alguns pontos que explicam essa performance com o objetivo de dar a ver a plataforma, no sentido adotado por Tarleton Gillespie (2010). As principais operações que integram a arquitetura do software da WBM são as seguintes:

1. Adição de tags para corrigir uma URL
2. Modificação de tags que podem provocar a mudança do pano de fundo de uma página
3. Inserção de um script em java que é executado no navegador do usuário, após a página ser baixada e que é responsável pelo redirecionamento de todos os endereços (URLs) daquela página para a WBM.

Estes protocolos fazem com que ao pesquisar uma página em determinado navegador parte do conteúdo em flash ou java sejam frequentemente redirecionados para o website atual. Algumas vezes todo o documento pode ser substituído pelo conteúdo online ou sequestrado, como descreve Tofel: *javascript in the original document occasionally will redirect the browser to a page on the live web, in effect hijacking the Wayback replay session* (2007;8). Quando uma sessão é capturada pela rede atual, perde-se ao mesmo tempo o acesso ao recurso da linha do tempo em que apareciam os conteúdos capturados pela WBM. Neste caso o pesquisador não possui recursos para manipular o acesso ao documento. Mas, há a possibilidade de entrar em contato com o Internet Archive e solicitar a recaptura dos dados, porém não é sempre que esta pode ser feita.

As páginas arquivadas carregam em seu próprio código os rastros da operação de arquivamento. Para Niels Brugger (2009) os arquivos da web produzem uma mediação relevante na criação de um novo tipo de fonte: a página arquivada da web. Este objeto distingue-se da página ainda em circulação que pode ser o objeto de estudo de outros pesquisadores. De acordo com o historiador dinamarquês:

“the archived website differs significantly from the website on the live web as well as from other known archived media types (by ‘live web’ it is meant ‘what is on the web today’). (...)”

No matter how an archived website has been created and no matter which archive it is found in, the website historian cannot expect it to be an identical copy on the scale of 1:1 of what was actually on the live web

at a given time. This is because, on the one hand, the archived website is an actively created and subjective reconstruction, and on the other hand, it is almost always deficient.” (2009:125)

Deste modo, se por um lado na web as páginas parecem se materializar quase instantaneamente na tela do computador, quando o servidor está em um dia bom, na interface da Way-back Machine o usuário é confrontado com vários ruídos que podem dar a ver as estruturas da almejada invisibilidade das operações entre os cliques, uma característica comum na navegação na web (GALLOWAY, 2012). A navegação no passado da web carrega o desgaste do tempo inscrito na própria mídia do arquivo.

3.2.3 *O curto-circuito da memória em dados digitais*

“According to some estimates, by the end of this century the total amount of computer memory in the world will be greater than the total amount of information in the world. The curves will intersect. At this point the demand for storage space will give way to a demand for information. By that time historians will look back with a smile on our current worries about information overload.” (MEDIAMATIC, 1994)

De que maneira podemos relacionar a estratégia de preservação representada pelo Internet Archive com as conceituações em voga sobre a memória digital. Na verdade entendemos o Internet Archive e as formas a partir dali engendradas de instauração de práticas de preservação, modelos e formatos, como um elo central no cenário de uma cultura de memória marcada pelas criptografias, ou, como estamos procurando conceituar, um enredamento de dados que se faz a partir de uma sobreposição de agenciamentos tecnológicos e culturais.

A pergunta que podemos fazer é: qual era a proposta de arquivamento contida no Internet Archive? Podemos observar as semelhanças entre o discurso da instituição com a linha argumentativa adotada pela UNESCO ao lançar o Documento pela Preservação do Patrimônio Digital⁵⁸. Isto é, ambos apontam, com certo alarme, para o fato de que os dados nascidos e postados na web não estavam sendo incorporados pelas instituições culturais, colocando em cheque a própria transmissão do legado cultural de uma geração para outra. Esta questão será retomada mais adiante, no momento nos ateremos a nossa descrição do maior arquivo da web.

⁵⁸ Como já mencionamos no primeiro capítulo a Carta de preservação da UNESCO chamava atenção justamente para o risco de perda dos dados nascidos digitais, e, por conseguinte, da herança cultural ali registrada.

A infraestrutura computacional do arquivo faz emergir um sistema que ordena as páginas não por assunto ou ordem alfabética. De acordo com Wolfgang Ernest trata-se de uma ruptura no regime de arquivo:

“the traditional architecture of the archive has been based on classifying records by inventories; this is now being supplemented or even replaced by order in variation and fluctuation, that is, dynamics access. This ‘archive’ is no longer simply a passive storage space but becomes generative itself in algorithmically ruled processuality.” (2013:29)

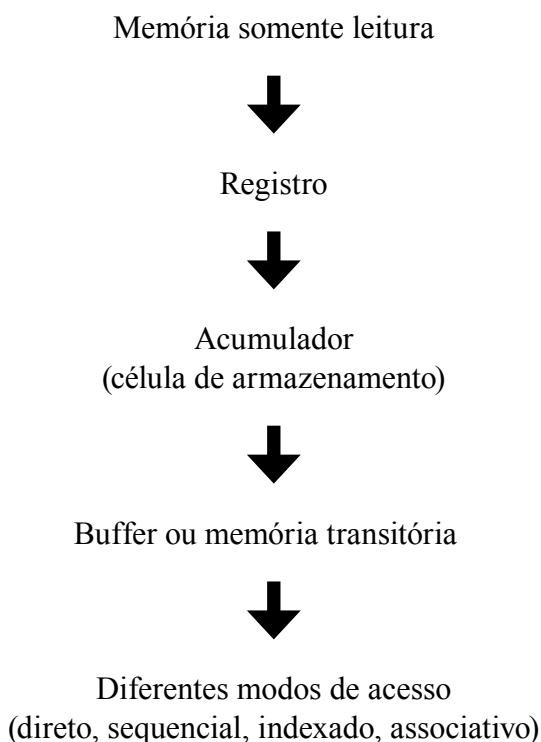
O sistema digital de armazenamento provoca também uma nova estética do arquivo, que passa a transmitir uma memória dinâmica e que implica uma reescrita para ser acessada (ERNEST: 2013;121). O arquivo constituído em código binário é necessariamente reescrito a cada operação de acesso, em cada browser, em cada tela. Isto se torna muito claro no caso do Internet Archive que arquiva páginas da web e está, ao mesmo tempo, imerso na transmissão de protocolos da rede. É válido lembrar que nem toda mídia digital é necessariamente reescrita ao acessar, o CD-ROM, por exemplo, é uma forma de preservação que contém um grau bem maior de fixidez embora possa igualmente manifestar algumas variações no acesso. Ainda, de acordo com Ernest:

“In the age of digitalizability, that is, the option to storage all kinds of information, a paradoxical phenomenon appears: Cyberspace has no memory.

Cyberspace is not even a space, but rather a topological configuration. That is why the metaphorical application of the Renaissance *ars memoriae* to Internet memory is a misapplication. There are no *lieux de memoire*, rather there are address. In the internet, the address structure of communication and the address structure of archival holdings merge into one. From place to pure address: traditionally, only what has been stored can be located and vice-versa.” (1999:199)

Impossível negar a emergência de uma nova estética do arquivo, porém pensamos que esta estética vai além do endereço. Em uma compilação mais recente de seu artigo, *Archives in Transition* (2013), Wolfgang Ernest argumenta que há uma hierarquia técnica nesta estética:

Imagem 27 - Esquema de hierarquia técnica na nova estética do arquivo proposto por Wolfgang Ernest



Esquema proposto por Wolfgang Ernest (2013)

Embora me pareça difícil uma aplicação analítica deste modelo, considerei que as distinções propostas indicam uma maior complexidade na apreciação do fenômeno. Na verdade, a citação anterior, de 1999, não consegue abranger a variedade de nuances que a web oferece nos dias atuais. Ao elencar de uma maneira arqueológica as camadas de composição de uma memória digital, Ernest desloca o argumento inicialmente proposto e provoca novas reflexões.

Partindo destas pistas nos interrogamos sobre como o Internet Archive poderia ser entendido por meio deste olhar da arqueologia das mídias. Isto é, de que modo a plataforma do arquivo, tal qual pensado por Brewster Kahle, possui uma dinâmica que romperia com a continuidade das práticas mnemônicas? De um lado identificamos os saltos temporais como uma performance já estabelecida nos códigos, porém não necessariamente programada. Assim, se nas linhas da programação da página arquivada não fizerem menção, por exemplo, ao atributo *refresh*, que usualmente recarrega a página após trinta minutos, a página arquivada poderá ser redirecionada para a web atual⁵⁹. Ao mesmo tempo, como já apontamos, toda a sessão pode ser

59 No Blog Web Science and Digital Libraries da Old Dominion University, Justin F. Brunelle fornece exemplos detalhados de como estas operações não são exclusivas do Internet Archive, sendo bastante comum em outros arquivos da web. Brunelle nomeia tais páginas como zombies, ou seja, as páginas

sequestrada por um script em *java* que redireciona do arquivo para a web atual. Tal característica só pode ser compreendida a partir de uma perspectiva que compreenda a materialidade deste arquivo, ou seja, o fato de seu conteúdo possuir qualidades técnicas semelhantes ao suporte de armazenamento. Forçando a analogia seria como dizer que os manuscritos e as estantes fossem feitos todos de um só material e formassem uma amálgama sem uma classificação hierárquica.

A gigantesca acumulação de dados e a relativa inacessibilidade do acervo não seriam os aspectos mais bizarros na análise de Wendy Chun (2006). O mais estranho neste modelo de arquivamento seria o fato do arquivo ser igualmente um arquivo do arquivo: “the Internet Wayback Machine’s greatest oddity, however, stems from its recursive nature; it and its archives (the IWM diligently archives itself) are themselves included among the objects of its archive” (2006:168). Na verdade, poderíamos pensar que todo arquivo, por meio de suas ferramentas de pesquisa, está igualmente se autoarquivando, porém tal operação se manifesta em cada registro o que faz com que nas transferências de dados ou no reprocessamento dos códigos o conteúdo do arquivo pode se tornar diverso da versão anterior. Neste sentido, concordamos com a argumentação da autora que propõe que no sistema de armazenamento da web em geral o que perdura é um contínuo e diligente processo de recuperação de dados. Os dados em si que circulam na rede tem como característica central a possibilidade de serem reescritos a cada processamento, mas são os e constituem sistemas diferenciados que compõem os modos variados de desaparecimento e reaparecimento das páginas da web⁶⁰.

Para tentar desvendar o enigma deste gigantesco arquivo, que se apresenta como biblioteca e se propõe a exibir artefatos culturais, optamos por partir das diferenças e semelhanças com as tradicionais instituições de memória. Por instituições de memória estamos nos referindo a uma recente nomenclatura que abrange museus, bibliotecas e arquivos, sob uma só denominação. De acordo com Helena Robison (2012) a utilização deste termo pode ser verificada a partir do final da década de 1990 e coincide com a ampla utilização das tecnologias digitais e do banco de dados por estas instituições. Em suas palavras: “initially, the use of term memory institutions to describe libraries, archives and museums was linked to the new possibilities opened up by the advent of the Internet and interoperable databases that could facilitate seamless access to collections information” (2012:415). Neste sentido é que pensamos situar a criação do Internet

mortas e vivas. Brunelle, Justin F. *Zombies in the archive*. (2012). Disponível em: <http://ws-dl.blogspot.nl/2012/10/2012-10-10-zombies-in-archives.html>

60 ‘the enduring ephemeral – a battle of diligence between the passing and the repetitive – also characterizes content. The internet may be available 24/7, but specific content may not. Further, if things constantly disappear, they also reappear, often to the chagrin of those trying to erase data’ (CHUN, 2006:167)

Archive em 1996, apontando suas especificidades em um contexto de convergência e adoção de novos papéis mesmo pelas instituições já estabelecidas.

3.2.4 *O Internet Archive é um arquivo?*

Para tentar responder a esta pergunta é necessário primeiramente explicitar o que estamos aqui chamando de arquivo e em que medida a proposta e as práticas de registro e coleta do Internet Archive se adéquam a este modelo. Uma premissa comum às instituições arquivísticas, que se expandiram sobretudo no Ocidente no século XIX, associadas a um Estado Nacional, é a função de guarda de documentação do aparelho burocrático (Featherstone, 2006). Neste sentido de arquivo, que Ernest chama de arquivo clássico, não poderíamos incluir a proposta do Internet Archive. Mas, como se poderia definir um arquivo no século XXI? Eis uma definição que consta em um clássico da arquivologia do Brasil:

“ Arquivo – É a acumulação ordenada dos documentos, em sua maioria textuais, criados por uma instituição ou pessoa, no curso de sua atividade, e preservados para a consecução de seus objetivos, visando a utilidade que poderão oferecer no futuro” (Paes, 2013:16)

Neste caso, a maior dificuldade de pensar o Internet Archive a partir desse conceito de amplo escopo está, principalmente, na questão do ordenamento da documentação. Embora exista uma arquitetura da informação ali depositada e esta deva ser alvo de nossa atenção, não podemos dizer que exista uma classificação hierárquica nem tampouco qualquer ordenação realizada por arquivistas. A coleta de metadados se dá de modo automatizado e por princípio não há descarte da documentação. Não há, portanto, nenhuma tarefa de seleção (embora seja excluída qualquer página que o autor solicite via formulário ou via código de exclusão automática).

Como já sinalizado pensamos que o Internet Archive pode ser visto não como uma instituição arquivistas tradicionais, que se fortaleceram sobretudo no século XIX associadas ao Estado e a disciplina histórica. Também não poderia ser entendido como arquivo, no sentido amplo adotado por Michel Foucault, em *Arqueologia do Saber*, que associa o arquivo a uma lei geral, ou um sistema de enunciados (1995:148). Somente poderíamos pensá-lo como arquivo se adotarmos a conceituação proposta por Aleida Assmann (2011), que o coloca em oposição ao canon. Neste sentido específico de manancial, de repositório reserva é possível pensar no total dos dados preservados nos servidores da Califórnia e do Egito.

No entanto ainda assim as especificidades deste projeto não caberia tão somente no modelo conceitual proposto por Assmann (2011). Isto porque a plataforma pública do arquivo, ou seja, sua página na web, exerce uma atividade de caráter participativo, engajando-se em assuntos do presente, como as eleições. Uma das primeira coleções disponibilizadas era justamente de páginas políticas norte-americanas e que foram expostas numa exibição no Museu Smithsonian. Hoje, o Internet Archive possui uma conta no twitter onde divulga regularmente conteúdos interessantes de seu acervo. Há uma sobreposição nas atividades desempenhadas que mesclam as antigas funções do museu, do arquivo e da biblioteca se pensarmos no Internet Archive enquanto instituição. No próximo capítulo aprofundaremos alguns destes pontos ao tratarmos da elaboração da coleção Buscas.br, a partir dos arquivos ARC e WARC acessados via WayBack Machine.

4. BUSCAS.BR: UMA EXPERIMENTAÇÃO NO ARQUIVAMENTO DE PÁGINAS DA WEB BRASILEIRA

O ato de compor uma coleção está no cerne das operações mais básicas da pesquisa histórica ou social que busca construir um corpus documental com o propósito de investigar fatos, grupos, técnicas etc. Nosso estudo, portanto, insere-se numa longa tradição de busca de evidências para a construção de uma possível narrativa sobre a história da internet no Brasil, ao mesmo tempo em que se propõe a entender como se pode investigar tal história através de fontes nascidas digitais. Neste sentido, seguimos a perspectiva de Michel de Certeau de que “tudo começa com o ato de separar, de reunir, de transformar em ‘documentos’ certos objetos distribuídos de outra maneira” (CERTEAU, 1982:81). Nesta pesquisa elegemos a própria coleta dos dados como uma estratégia no sentido de tentar responder a nossa pergunta inicial: como se pesquisa o passado da web a partir das fontes nascidas digitais? Mais do que propor, a partir dos documentos, contar uma história, a aposta que se desenhcou foi a de narrar as possibilidades disponíveis e inventadas, relativas aos vestígios digitais dos primórdios da web brasileira. Ao mesmo tempo o próprio ato de colecionar, sejam documentos ou objetos, implica na construção de sentidos. Desse modo, a proposta de selecionar algumas centenas de páginas dentre as milhões contidas no Internet Archive implica numa operação que irá dotar de valor a experiência, tornando as ruínas digitais objeto de uma coleção.

Ao mencionar o processo de criação de uma coleção, não podemos deixar de mencionar a obra do historiador Krystoff Pomian, que propõe o termo ‘semióforo’ para os objetos dos museus, uma vez que eles fazem a mediação entre o visível e o invisível (1990:5). Porém, embora de uma forma bastante geral possamos relacionar qualquer ato de colecionar a uma operação valorativa, é importante notar algumas distinções relevantes entre as relíquias, os tesouros ou objetos funerais arrolados por Pomian em sua obra clássica e os objetos digitais, alvo de nossa pesquisa. Isto porque, para além da dimensão material dos objetos selecionados para integrarem as coleções citadas pelo autor, o que mais diferencia estas coleções é o propósito da exibição. O sentido universal da coleção estaria, portanto, na retirada dos objetos do fluxo de utilidade social (comercial, sagrado, etc) e a realocação em outro circuito com objetivos de exposição (POMIAN, 1990:174). Se tentarmos estender o alcance de tais considerações aos objetos digitais, um problema fundamental se apresenta: o objeto digital não é retirado do seu fluxo; ele é apenas reproduzido em uma outra plataforma, que certamente implica usos e significados diversos. Como então conceituar o colecionismo digital?

Em primeiro lugar, conforme apontamos no primeiro capítulo, há práticas bastante diversificadas referentes às operações de colecionar, armazenar e exibir os objetos digitais. Algumas dessas práticas implicam na sacralização do objetos, enquanto outras operam em outros regimes. Se podemos identificar uma tendência na convergência dos papéis dos museus, bibliotecas e arquivos no que diz respeito às ações de preservação da memória cultural (KATELAR, 2001), há, por outro lado, que se tentar discernir as diferenças ainda existentes nas esferas de ação próprias a cada instituição de memória (ROBINSON, 2012). Neste sentido, a diferença que estamos apontando está para além do nítido contraste que se faz notar entre a fisicalidade do objeto resguardado por um museu e a expressão numérica de um registro digital. Buscando dar conta da heterogeneidade de um conjunto de coleções digitais, nos deparamos com a importante distinção proposta por Aleida Assman (2010) entre os conceitos de Canon e de arquivo.

Aleida Assman estabelece uma dinâmica bem distinta entre as esferas do canon, um espaço de produção de memória cultural ativa, e o arquivo como uma esfera onde seria possível identificar um passivo do legado cultural⁶². Um exemplo citado pela autora procura contrastar dois espaços distintos em um museu: as salas de exibição funcionariam como canon e a reserva seria o arquivo. A proposição de tais conceitos visa identificar as complexas práticas envolvendo a produção e transmissão da memória cultural. Neste sentido, Assman considera que:

“the institutions of active memory preserve the *past as present* while the institutions of passive memory preserve the *past as past*. The tension between the pastness of the past and its presence is an important key to understanding the dynamics of cultural memory.” (ASSMAN, 2010:98)

A escolha de criar uma coleção de páginas da web nos coloca, primeiramente, diante de uma esfera passiva da memória documental, uma vez que o acervo do Internet Archive não é facilmente acessado, nem mesmo por pesquisadores da área de história da rede mundial de computadores. O acesso às ruínas das páginas da web brasileira preservadas exige, portanto, uma atitude ativa. Porém tal processo não implica em um deslocamento para o canon. Como esperamos deixar claro, logo em breve, nos parece que ao propor esta coleção e estabelecer com isso novas possibilidades de acesso a uma pequena porção do passado da web brasileira, estamos ainda na esfera do arquivo, tal qual definido por Assman. Há, por outro lado, dentro da dimensão do arquivo, uma série de fluxos de uma instância (ou de uma camada) a outra. Tais fluxos podem ser aproximados de certos processos relativos às operações de colecionar e

⁶² A matriz desses conceitos estaria na obra do historiador Jacob Buckhardt, que propôs os conceitos de mensagem e traços para discernir graus distintos de atuação da memória cultural (ASSMAN, 2010:99).

tornar visível o invisível, descritos por Pomian, entre outros autores. Voltaremos a esta reflexão ao final deste capítulo, pois a descrição analítica que se segue servirá de evidência para que possamos avançar na discussão sobre a adequação dos conceitos e, ao mesmo tempo, retomar a noção de criptografias da memória. Antes de iniciar a discussão é importante ainda esclarecer um último ponto. Embora todos os esforços para a realização da coleção estejam voltados para proporcionar o acesso livre à documentação, buscando dar maior visibilidade aos vestígios digitais, pela própria natureza dos dados, é provável que apenas um grupo restrito de pesquisadores faça uso do trabalho aqui apresentado.

4.1 Uma coleção de páginas?

A proposta de construir uma coleção teve como objetivo principal a investigação sobre os modos e possibilidades envolvidas no arquivamento e no acesso ao passado da web brasileira. Porém tal proposta implicou também em escolhas relativas ao conteúdo a ser armazenado, assim como ao tipo de amostragem requerido. Como já mencionamos, ao investigarmos as ações das instituições de memória no Brasil pudemos constatar a ausência de experiências relevantes de arquivamento da web. Antes de iniciar o processo de elaboração da coleção, realizamos um segundo mapeamento para averiguar a possibilidade de pesquisa de fontes brasileiras em outras instituições e encontramos a coleção “Brazilian Presidential Election 2010 Web Archive” na Biblioteca do Congresso dos Estados Unidos. Fizemos uma sondagem nesta coleção, cogitando utilizá-la como material de pesquisa e, embora ela não tenha se tornado nosso objeto, consideramos importante falar deste caso raro de preservação da web brasileira que foi nosso primeiro ponto de parada.

A coleção “Brazilian Presidential Election 2010 Web Archive” foi criada pelo programa Minerva da referida biblioteca que, desde 2001, coordena a elaboração de coleções temáticas de páginas da web. Tendo como propósito mapear as discussões em torno das eleições presidenciais de 2010, a coleção foi construída através de varreduras realizadas entre setembro de 2010 e janeiro de 2011, em 48 websites selecionados por uma equipe que envolveu o escritório da Biblioteca do Congresso no Rio de Janeiro.

Imagem 28 - Arquivo sobre as eleição presidencial no Brasil do website *Library of Congress*



Fonte: Library of Congress (2014)

O pesquisador pode acessar a “Brazilian Presidential Election 2010 Web Archive” de várias formas. Há um catálogo por assunto com 51 itens, um índice onomástico com 42 itens e uma lista dos endereços capturados contendo os 48 URLs. Além de possibilitar o acesso a esses instrumentos de pesquisa, ainda é possível fazer uma busca por palavras na base de dados das fichas catalográficas dos websites. Esta busca por palavras não realiza, no entanto, uma pesquisa diretamente nos textos arquivados das páginas, mas sim no um resumo do conteúdo arquivado, que inclui palavras-chave. O acesso é, portanto, bastante facilitado se compararmos, por exemplo, com o Internet Archive. Porém, a consulta fica restrita aos metadados, pois o conteúdo dos websites só pode ser acessado na sala de consulta da Biblioteca do Congresso dos EUA. Diferentemente das coleções das eleições norte-americanas, que podem ser consultadas online, esta coleção de páginas brasileiras encontra-se bloqueada.

Outro fator que dificulta o acesso é o fato de não haver metadados em português, nem mesmo o título. Por conta desta restrição, na minha pesquisa, só consegui chegar até a coleção ao buscar diretamente no website da Biblioteca do Congresso, embora já tivesse realizado inúmeras buscas anteriores em diferentes bases. A “Brazilian Presidential Election 2010 Web Archive” representa certamente um conjunto importante de fontes, mas, além das dificuldades de acesso, o seu escopo está restrito a um tema e temporalidade muito específicos, o que não atendia às inquietações da nossa pesquisa.

Se a “Brazilian Presidential Election 2010 Web Archive” nos parecia um terreno demasiado limitado, por outro lado, nos deparávamos com a dificuldade de delimitar um campo de estudo em meio à vastidão de informações que caracteriza a web. Toda pesquisa implica em escolhas, em recortes, mas ao lidar com um universo como a história da web, a tarefa de delimitar o objeto é uma etapa fundamental, pois como afirmam Fragozo, Recuero e Amaral:

“a internet é um universo de investigação particularmente difícil de recortar em função de sua escala (seus componentes contam-se aos milhões e bilhões), heterogeneidade (grande variação entre as unidades e os contextos) e dinamismo (todos os elementos são permanentemente passíveis de alteração e a configuração do conjunto se modifica a cada momento)” (2012:55)

Os autores estão se referindo principalmente à web atual, ou seja, à pesquisa realizada no presente. Mas a análise apresentada se aplica em grande medida ao nosso trabalho, que se propõe a investigar a web 1.0, pois o tipo de fonte a ser consultada é semelhante, muito embora as páginas arquivadas tenham especificidades que serão exploradas a seguir. A pergunta e o quebra-cabeça que inicialmente se apresentaram estavam relacionados ao quê investigar e com qual propósito. Não é novidade que a pergunta ao passado é sempre realizada com os pés bem fincados no presente. E esta preocupação com o presente foi o que certamente motivou o recorte da nossa pesquisa. O início do século XXI é marcado pela crescente aceleração do número de usuários da web e por uma expansão gigantesca do número de páginas. Ao mesmo tempo, forjou-se ao longo do processo de comercialização da rede um avanço de políticas monopolistas, sobretudo no que se refere à porta de entrada da rede - as ferramentas de buscas.

Diante deste novo fenômeno surgiram muitos estudos importantes que chegaram a caracterizar este período como “sociedade das buscas” (Society of the Query⁶³), na denominação de Geert Lovink (2012); “sociedade das ferramentas de busca” (search engine society), nos termos de Alex Halavais (2008) ou ainda a Googlelização (googlelization) para Siva Vaidhyanathanou (2011). Todos os autores chamam atenção para o papel que os engenhos ou ferramentas de busca desempenham na arquitetura social, política e cultural da rede mundial de computadores. A emergência de uma sociedade baseada nas ferramentas de busca, ou seja, onde as plataformas que dão o acesso às informações contidas na web exercem uma proeminência significativa, pôs em xeque o discurso utópico sobre a Web. Desde o início do século XXI ficava claro que o ideário do acesso democrático à informação, associado à Web, estava sendo desbancado pela vinculação entre artefatos técnicos estratégicos e monopólios comerciais.

Uma nova área de estudos foi aos poucos se formando e, a partir dos anos 2000, começaram a se delinear algumas linhas de pesquisa importantes, como aquela referente aos estudos

⁶³ O grupo de pesquisa *Society of the Query*, do Intitute of Network Cultures disponibiliza uma lista de referências importantes para o estudo das ferramentas de busca. Da listagem, destacamos a seguinte obra: IPPOLITA [coletivo], *The Dark Side of Google*. Translator: Patrice Riemens. Editorial support: Miriam Rasch. Publisher: Institute of Network Cultures, Amsterdam 2013. ISBN: 978-90-818575-6-7. Disponível online: <http://networkcultures.org/wpmu/query/resources/readings/>.

de usabilidade e design. Michel Zimmer elabora um balanço sobre os estudos recentes do tema e enumera algumas áreas a serem contempladas:

Future research must ensure continued progress in the multidisciplinary understanding of the design, use, and implications of Web search engines. Four research areas can quickly be identified that deserve particular attention: search engine bias, search engines as gatekeepers of information, values and ethics of search engines, and framing the legal constraints and obligations (2009:516).

Seguindo a trilha lançada por estes autores, percebemos que as ferramentas de busca constituem um nó importante da rede, uma vez que impõem uma determinada ordem aos milhões de dados da web. Passamos então a nos perguntar sobre como seria pesquisar e acessar a internet brasileira antes da era Google. Diante desta problemática, pensamos construir um corpus sobre as ferramentas de buscas brasileiras da web 1.0, período de expansão da rede mundial de computadores, cujos marcos cronológicos são a expansão da web comercial (1995) e o estouro da bolha da internet em 2001 (O'REILLY, 2002; STEVENSON, 2013).

Nossa escolha segue uma perspectiva que busca chamar atenção para a necessidade de uso das fontes digitais para o estudo da web, pois embora existam inúmeras pesquisas sobre o passado da web, a utilização dos arquivos digitais ainda é incipiente (BRUGGER, 2008).

De um lado, a experimentação com a coleção tinha que possibilitar uma investigação sobre como pesquisar o passado da web, mas, ao mesmo tempo, como uma outra face do espelho, nossa pergunta dirigia-se também a uma documentação nativa digital para responder às questões sobre o processo histórico que envolveu a conformação de um processo global de ascensão do Google. Hoje, no Brasil, mais de 90% das consultas à internet iniciam-se na página do Google⁶⁴, enquanto que, em 1998, quando ocorreu uma das primeiras medições das ferramentas de busca no país, as principais páginas eram o Altavista, o Cadê e o Bol/Uol (este funcionava com tecnologia do Todo.Br). Embora nossa pesquisa não possa dar conta de explicar as razões econômicas relacionadas a este processo, a coleção pode expor/dar a ver vestígios que auxiliem na elaboração de estudos futuros. Ao mesmo tempo, nos interessa dar a ver como era a navegação na web a partir de 1997.

A proposta da coleção procura criar um corpus com as ferramentas de busca brasileiras, aprofundando a coleta nas categorias relacionadas a História e Cultura. A escolha desta temática constituiu-se como mais uma estratégia para o desdobramento de nossa problemática.

64 Fonte: Olhar Digital, disponível em: <http://olhardigital.uol.com.br/noticia/38245/382> (acesso em 10/09/2014)

Uma vez que estamos investigando as criptografias da memória, entendidas como operações culturais mediadas por tecnologias da informação, consideramos que páginas voltadas para a cultura e a história poderiam oferecer pistas sobre a relação estabelecida pelos usuários com estes temas, muito embora o conteúdo das páginas arquivadas não fossem o foco principal da pesquisa. Além disso, consideramos importante a escolha de um tópico com o qual tivéssemos a devida familiaridade para poder contribuir em futuras inclusões de metadados e na criação de sub-coleções. Foram selecionados inicialmente quatro principais ferramentas de busca:

Tabela 2 - As quatro principais ferramentas de busca

Nome	Criação		Ano de venda/ extinção	Observações
	Ano	Local		
Uol	1996	São Paulo	Em funcionamento	Criado pelo grupo Folha da Manhã, como provedor e portal de conteúdos, o UOL passou a ter uma ferramenta inicialmente como catálogo e posteriormente, utilizando a tecnologia do TodoBr
Cadê?	1995	Rio de Janeiro	1999 para o Grupo Terra e em 2001 para o Yahoo!	Criado por Gustavo Viberti e Fábio Oliveira.
Radix	1997	Recife	2001, para o banco Oportunity	Criado a partir do departamento de Informática da UFPE.
TodoBr	1999	Belo Horizonte	2005 para o Google	Criado a partir do departamento de Informática da UFMG.

4.1.1 Como fazer uma coleção de páginas?

Relato de uma incursão no emaranhado de dígitos

Definida a temática, surge outra questão: como poderíamos coletar as páginas passadas da web? A seguir, contamos brevemente o percurso investigativo que conduziu até o projeto

de parceria com o Archive-It, a via escolhida para a realização da coleção. Esta resolução veio após inúmeras tentativas de coletar as páginas e organizar os arquivos no meu próprio computador, processo que demandou um significativo dispêndio de tempo na busca de informações em fóruns online e em órgãos de preservação digital, como o International Internet Preservation Consortium (IIPC).

Ao longo da pesquisa pude verificar que a proposta de alguns grupos de pesquisadores, sobretudo vinculados à Digital History norteamericana⁶⁵, de fomentar o aprendizado de linguagens computacionais é importante, mas provavelmente os futuros historiadores devem começar a formação desde a graduação. Mesmo tendo me aventurado em algumas interfaces como o SEARS (Software Environment for the Advancement of Scholarly Research) e seguido os cursos online da *Code Academy* percebi, após sucessivos percalços, que a empreitada era longa e exigiria abrir uma nova frente de estudos. Além disso, a proposta era realizar um trabalho que pudesse ser útil a outros pesquisadores, o que requer máquinas de um porte muito superior ao de um computador pessoal, de modo a garantir o acesso de longo prazo. Outro fator que impossibilitaria o desenvolvimento de tal acervo num computador pessoal seria a capacidade de armazenamento/processamento, pois muito embora toda a coleção possa ser armazenada em um pequeno disco externo, a capacidade de processamento exigida na coleta é grande. O trabalho com os *crawlers* (programas que capturam automaticamente as páginas listadas), por exemplo, mostrou-se bastante complexo. O *crawler*⁶⁶ ou robô de busca é um software programado para seguir determinada linha de busca e armazenar as páginas visitadas. Certos *crawlers* como, por exemplo, o programa SEARS (Software Environment for the Advancement of Scholarly Research⁶⁷), exigem um servidor para processar os dados e só seria possível utilizá-lo em parceria com alguma instituição. Nas minhas experiências testando a possibilidade de uso de alguns *crawlers* me deparei com várias quedas do sistema ou com um tempo tão longo para fazer o download dos dados que tornava inviável o prosseguimento da tarefa.

Participei de experiências que traziam outra visão sobre o que o pesquisador deve (ou pode) realmente saber e realizar, como os grupos interdisciplinares onde cada um contribuía

65 Ver, por exemplo, a proposta do projeto Programming Historian, desenvolvido por Willian Turkel, Adam Crymble e Fred Gibbs, cujo objetivo é possibilitar o aprendizado de diversas linguagens computacionais por historiadores. Disponível em: <http://programminghistorian.org/> (acesso em 10/04/2014).

66 Em Portugal se utiliza o termo batedor para designar tal software, porém todas as fontes brasileiras consultadas, assim como a fala de todos os entrevistados, mencionavam apenas o termo original em inglês, por isso optamos também por manter a palavra *crawler*.

67 Até 2012, o programa não possuía ferramentas compatíveis com a língua portuguesa. Disponível em: <http://www.lis.illinois.edu/research/projects/software-environment-advancement-scholarly-research-sears> (acesso em 10/04/2014).

com a sua área de especialização. Mas, tentar fazer a coleta de páginas por mim mesma/de forma autônoma, serviu, ao menos, para demonstrar a impossibilidade de desempenhar uma tarefa de modo isolado e sem suporte técnico. Testei, por exemplo, o programa WAIL (Web Archiving Integration Layer), que foi criado em 2013 por Matt Kelly, um estudante de pós-graduação da Old Dominion University. O WAIL possibilita a coleta e a visualização por meio do WayBack Machine, porém oferece poucos recursos para manejo dos metadados⁶⁸. Embora a impossibilidade de seguir com a experiência tenha se mostrado patente frente aos obstáculos surgidos, certamente o aprendizado realizado ao longo de todo este percurso foi incorporado ao trabalho, uma vez que os conhecimentos construídos possibilitaram um maior entendimento das funcionalidades dos arquivos nascidos digitais.

Buscando sondar outras possibilidades para a pesquisa, me inscrevi em um treinamento online do Internet Archive para aprender sobre as especificidades do Archive It. A partir deste workshop via web tive a possibilidade de conhecer as características técnicas e os cenários de utilização desta plataforma. Percebi que esta seria a via adequada para explorar de modo mais próximo a pesquisa sobre o passado da web brasileira, sem a necessidade de adquirir todo um corpo de conhecimentos em linguagens de programação. O Archive-It oferece serviços pagos, a maior parte deles para bibliotecas ou instituições governamentais, porém também aceita parcerias com pequenos projetos como o nosso. Em janeiro de 2013, o Archive-It possuía um total de 238 instituições parceiras, a maioria nos Estados Unidos. Além disso, 15 países possuem coleções digitais no repositório do Archive-It (BRAGG, HANNA, 2013:2).

O projeto Archive-It foi criado em 2006 e inicialmente contava com apenas 13 instituições parceiras. O maior objetivo do projeto era justamente prover as instituições de memória com ferramentas de coleta já criadas pelo Internet Archive, no caso o crawler Heritrix, assim como desenvolver uma plataforma que possibilitasse a curadoria e o acesso às diferentes coleções. Neste sentido, a criação deste programa procurava atender à missão do Internet Archive, conforme mencionamos no capítulo anterior, de possibilitar o acesso e a preservação das páginas da web. Ao mesmo tempo, o desenvolvimento do Archive-It expõe as limitações que a interface da Wayback Machine colocava para usuários finais e também para as instituições. Assim o projeto procura suprir as necessidades de pesquisa e de uma interface com flexibilidade suficiente para atender demandas distintas. De acordo como o relatório “The Web Archive Cycle Model”, elaborado a partir de pesquisas junto às instituições parceiras do Archive-It, fica explícita esta necessidade de novas ferramentas, uma vez que:

68 WAIL é um programa aberto e está disponível no website do seu criador onde também se encontra um vídeo com explicações básicas sobre sua utilização. Disponível em: <http://matkelly.com/wail/> (acesso em 10/04/2014)

“Despite growth in the number of web archiving programs, many institutions still struggle with developing best practices and methodologies to accomplish their goals. This difficulty partially stems from constantly evolving web technology, which can make it difficult to archive certain types of content effectively” (Archive It Team, 2013:1).

Na tabela abaixo procurei apontar, de modo resumido, as diferentes características do processo de armazenamento e acesso aos dados do Internet Archive e do Archive-It.

Tabela 3 - Diferentes características do processo de armazenamento e acesso aos dados dos websites *Internet Archive* e *Archive-It*

Internet Archive	Archive-it
Uma coleção de amplo espectro	Coleções com temáticas variadas
Varredura de snapshots a cada dois meses (disponível online após 6 meses)	Controle das instituições sobre o escopo e frequência das varreduras
Não há ferramenta de busca por palavras	Busca por palavras disponível em toda a coleção
Acesso automatizado via Wayback Machine	Possibilidade de suporte técnico
Não há possibilidade de catalogação de metadados	Metadados disponibilizados e indexados
Acesso público a todos os itens	O nível de acesso é variável

Fonte: Internet Archive. Documentação de suporte do Archive-It (2013).

O Archive-It possibilita a criação de coleções temáticas e, mais importante, permite o acesso aos dados através de buscas numa interface relativamente amigável. Isto é, o software possibilita que usuários possam pesquisar em uma base de dados previamente selecionada. Pensando na possibilidade de desenvolver uma coleção utilizando esta plataforma, escrevi um pequeno projeto, que detalharemos logo em seguida, resumindo a minha proposta e enviei um e-mail para a coordenadora do curso online, Lori Donavon, em setembro de 2013. A resposta foi muito cordial, tendo sido encaminhada para a diretora do Archive-It, Kristine Hanna, que me colocou algumas

questões. A partir daí trocamos algumas mensagens e depois combinamos uma entrevista via skype. Desde a primeira conversa me chamou atenção a curiosidade da interlocutora sobre as iniciativas de arquivamento da web no Brasil e estava claro o interesse em estabelecer uma parceria por parte do Internet Archive. Ao mesmo tempo, já nestas primeiras interlocuções foram explicitadas as possíveis dificuldades. Embora a coleção proposta, como veremos a seguir, seja de escala reduzida, os limites e desafios são significativos. Um dos aspectos que mais me chamou atenção diz respeito ao tempo longo necessário para capturar e ter acesso aos dados, que vai justamente de encontro às percepções correntes sobre a disponibilidade dos dados e à suposta velocidade em acessá-los. Na tentativa de tentar tornar o processo mais rápido, perguntei se seria possível eu mesma realizar a captura de dados. No seguinte trecho, em uma mensagem eletrônica de 10 de outubro, citado com autorização, Kristine Hanna afirma o seguinte:

“This is a very interesting project and we would love to work with you on the Brazilian Historical Web Sphere. Ideally for us we would start in December or January.

Also, in talking to the team here, we think it is more realistic for the Archive-It team to initially take the lead on creating and managing the collection; with you taking the lead on QA [Quality Assurance] with the two QA tools that we have available in the service - which we would provide training on. The QA would assist us in knowing what did not get captured, what got captured incompletely and so on.

We would still be working collaboratively with you and within the parameters of your project (as you want - we'd capture the content, name the collection, provide access etc).

And then maybe down the road as the crawling continues we can turn things over to you. How does that sound? (HANNA, 2013b)

Em dezembro de 2013 foi iniciada a parceria para construção da coleção, e a minha responsabilidade, como se depreende da mensagem acima, estava na curadoria dos websites, inclusão dos metadados e controle de qualidade. Houve, neste momento inicial, longos intervalos de tempo em que o trabalho foi suspenso devido a problemas técnicos na coleta e transferências do dados⁶⁹. O processo de controle de qualidade das capturas, que no caso da nossa coleção era

⁶⁹ Em um e-mail de 28 de maio de 2014, de Sylvie Rollason-Cass, fui informada que os engenheiros da equipe estavam tendo problemas na captura e transferência dos dados: Apologies for the delayed reply. (...)Also, I wanted to give you a quick update on full text search for your collection - There was an issue with the indexes of the transferred data which our engineers are continuing to work on. I'm sorry for the delay on this and I'm really hoping that it's something they will be able to fix in the next few weeks" (citado com autorização).

na realidade uma transferência do material previamente arquivado, demonstrou ser um trabalho que demandava mais tempo do que imaginado. Pude então compreender as razões pelas quais muitas instituições que mantêm parcerias com o Archive-It contam em sua maioria com dois funcionários (arquivistas ou bibliotecários) responsáveis pela coleção (ARQUIVE-IT TEAM: 2013). Na verdade, o processo de controle da documentação é parte fundamental na constituição da coleção e infelizmente não é um caminho muito pavimentado, como se pode perceber na seguinte passagem:

“After institutions capture data from their desired sites, they review what they archived and assess its quality and completeness (...). This can be done through reports generated by crawlers or by clicking through the archived websites themselves by way of an access tool like the Way-back software. The process of web archiving can include trial and error. Like most aspects of web archiving, no single best practice for Quality Assurance (QA) has emerged among institutions that archive the web.”(BRAGG, HANNA, 2013:27)

Assim, cada avanço na construção da coleção ocorreu após muitas tentativas e erros. A listagem base dos endereços que serviram de *seeds* foi atualizada várias vezes e ainda assim há problemas, que conforme veremos só serão enfrentados na fase de adição dos metadados. Uma das dificuldades estava, ao que tudo indica, na venda das empresas responsáveis pelas ferramentas de busca e no consequente redirecionamento dos endereços, o que teria gerado algumas quebras no sistema e impossibilitado a transferência dos dados. Ainda assim, em junho de 2014 a coleção estava relativamente pronta e em agosto pudemos colocá-la no modo público para teste. Eis a página inicial da coleção:

Imagem 29 - Coleção *Brazilian Web Engines (1997-2013)*

The screenshot shows the Archive-It interface for the 'Brazilian Web Engines (1997-2013)' collection. At the top, there is a navigation menu with 'HOME', 'EXPLORE', 'LEARN MORE', and 'CONTACT US'. The main content area features a search bar with 'buscas.br' entered. Below the search bar, there is a description of the collection, including the collector's name (Universidade Federal do Estado do Rio de Janeiro) and the date it was archived (Jan, 2014). The description states that the collection preserves captures from Brazilian web engines and was curated by the Internet Archive. It also mentions that the collection focuses on the History and Culture categories of the search engines. Below the description, there is a 'Narrow Your Results' section with a search box and a 'Search' button. The search results section shows 'Portuguese (4)' and a 'Search Page Text' button.

Fonte: Archive-It

O projeto da coleção teve como um dos objetivos facilitar o acesso a uma memória documental, nascida digital, porém em momento algum pensamos que este acesso seria universal. Ainda assim, importa notar as facilidades advindas a partir da sua concretização. Se antes a plataforma da Wayback Machine não permitia a busca por palavras, no Archive-It esta busca é possível, o que já facilita bastante a pesquisa. A busca avançada possibilita também a pesquisa a partir de marcos temporais, mas neste aspecto a plataforma possui uma falha. Embora a coleção possua material a partir de 1997, só é possível pesquisar na busca avançada a partir do ano 2000. Ao questionar esse atributo e perguntar se seria possível uma modificação na arquitetura desta forma de recuperar as informações, a equipe do Archive-It acolheu a demanda e ficou de enviar tais solicitações aos programadores como possível melhoramento nas próximas versões⁷⁰.

⁷⁰ Resposta recebida de Sylvie Rollason-Cass em um treinamento online em 9 de Julho de 2014.

Imagem 30 - Resultado da pesquisa por *História do Brasil* no website *Archive-It*

The screenshot shows the Archive-It search interface. On the left is the 'Advanced Search' panel with fields for 'Contains all of:', 'Exact phrase:' (containing 'História do Brasil'), 'Not containing:', and 'From the Host:' (containing 'ex. www.archive-it.org'). Below these are filters for 'Results per host:' (set to 1), 'File format:' (set to All formats), and 'Capture date range:' (From: January 2000 to January 2003). An 'Advanced Search' button and a 'Help with Search' link are at the bottom of the panel.

The main search results area is titled 'Page 1 of 9 (168 Total Results)' and includes a 'Next Page' button. The results are sorted by 'Best Match'. Three results are displayed:

- História do Brasil**: URL: <http://www.historiadobrasil.hpg.com.br:80/>. This text was captured 4 times between Apr 01, 2001 and Jul 20, 2001. [Show All Captures](#). História do Brasil História do Brasil Esta página apresenta todos os fatos que ocorreram no Brasil desde o seu descobrimento. Para que esse site seja visualizado com frame, clique aqui , ou então para acessar o site sem frames, clique aqui Content: text/html Size: 2.1 KB. [More Results from historiadobrasil.hpg.com.br](#)
- ClioNet - Rede Eletrônica da História do Brasil**: URL: <http://www.clionet.uff.br:80/>. This text was captured 9 times between Nov 22, 2002 and Jun 19, 2003. [Show All Captures](#). ClioNet - Rede Eletrônica da História do Brasil http://www.clionet.uff.br clionet@ah.uff.br... História OBJETIVO: A ClioNet, Rede Eletrônica de História do Brasil, através de suas malhas, tem por objetivo contribuir para a integração e o intercâmbio entre pesquisadores de História do Brasil, constituindo-se em um centro de referência virtual para a pesquisa em História do Brasil. HISTÓRICO... outros tipos de canais informais). Você é o visitante número MALHAS DA REDE ELETRÔNICA DE HISTÓRIA DO BRASIL - Revista Eletrônica de História do Brasil: rehb - Lista de Discussão de História do Brasil: hbrasil-1 - Clio Edições Eletrônicas: clioedel - Índice de História do Brasil na Web: hbrindex - Biblioteca Virtual de História do Brasil: bvhr - Índice de Revistas Brasileiras de História do... Content: text/html Size: 6.4 KB. [More Results from clionet.uff.br](#)
- Viagem pela História do Brasil**: URL: <http://www.historiadobrasil.com.br:80/>. This text was captured 8 times between Jan 16, 1999 and Mar 02, 2000. [Show All Captures](#). Viagem pela História do Brasil Apoio cultural: A aventura humana de montar uma civilização nos trópicos, contada com a tecnologia mais avançada, num livro de 352 páginas em cores e num CD-ROM com 1200 temas e mais de 2000 imagens. Uma história onde se cruzam povos, sonhos, guerras, cobiça, beleza e muitas novidades. A obra | Uma visão de Brasil | Viagem ... | Autores | Agenda e mídia | Cap.1 | Ajuda Webite criado por Maria Amélia de Azevedo e Flavio de Carvalho... Content: text/html Size: 4.9 KB. [More Results from historiadobrasil.com.br](#)

Fonte: Archive-It

Explicitar aqui esta falha técnica é importante para fazer ver como a interface do software está relacionada às práticas de arquivamento. Seguindo a trilha dos estudos de software, sobretudo os trabalhos já mencionados de Manovich (2001, 2008) e Fuller (2005), buscamos entender o porquê desta impossibilidade de configurar uma busca entre os anos de 1997 e 1999. O que poderíamos pensar desta especificidade? Uma primeira hipótese é de que isto esteja relacionado a uma atualização feita no próprio sistema do Internet Archive que, a partir de 2000, passou a arquivar as páginas em WARC e não mais em ARC. Tal mudança pode estar ligada a uma incompatibilidade técnica com a ferramenta de busca por palavras, porém estas são conjecturas não confirmadas pelo Archive-It. De todo modo, é interessante perceber o fato de que falha não se mostrou uma questão relevante para a equipe do Archive-It, deixando ver que o foco de maior interesse da plataforma tem recaído sobre um passado super recente, relacionado, em geral, às instituições que passaram a arquivar suas próprias páginas

a partir do ano 2000. O passado não tão recente, sobre o qual a nossa pesquisa se debruça, parece ter sido um pouco relegado na arquitetura do software. É preciso pontuar que na busca geral por palavras pode-se obter resultados dos anos de 1997 a 1999, porém sem a possibilidade de recuperação de tais dados em separado.

4.2 A arquitetura da coleção

Os pontos de partida (*seeds*) da coleção são as páginas iniciais de cada portal, que em si mesmas já oferecem uma quantidade relevante de informação sobre o estágio da web no final dos anos 1990. A partir desta primeira seleção, escolhemos também um segundo nível das categorias, que foram também incluídas como *seeds* (pontos de partida para o crawler). Neste segundo nível, selecionamos categorias aplicadas em cada ferramenta buscadora, nas seguintes áreas temáticas: cultura, educação, história, ciências humanas. Esta fase da pesquisa se revelou bastante desafiadora, uma vez que não encontramos uma bibliografia específica sobre o tema, nem do ponto de vista da história bastante recente das ferramentas de busca brasileiras, nem no que diz respeito ao uso de páginas da web como fonte principal em estudos experimentais no país⁷¹. Assim, seguimos o mote da equipe do Archive-It e avançamos com tentativas e erros.

Abaixo está a tela do Archive-It, com a listagem de *seeds* a partir dos quais pode-se consultar todas as páginas arquivadas.

⁷¹ Os estudos da área da Ciência da Computação envolvem problemáticas distintas, mas trouxeram algumas referências importantes (RIBEIRO, 1998).

Imagem 31 - Tela do Archive-It com a listagem de seeds a partir dos quais pode-se consultar todas as páginas arquivadas

The screenshot shows the Archive-It web interface. At the top, there is a navigation bar with links: Home, Collections, Crawls, Reports, Access, Help Documentation, and Submit a Question. Below this, the 'Access' section is active, featuring a 'Search' tab and a 'Wayback' tab. A search box is present with the placeholder text 'Specify an exact URL' and a 'Go' button. Below the search box, there is a message: 'Select a collection from the drop down menu below to browse seeds in that collection.' A dropdown menu is set to 'Brazilian Web Engines (1997-2013)'. Below the dropdown, a list of URLs is displayed, each with a small external link icon. The URLs include: <http://br.cade.yahoo.com/>, <http://busca.uol.com.br/>, <http://cade.com.br/>, <http://cade.search.yahoo.com/web/>, <http://mdi.uol.com.br/ciencia/historia/index.html>, <http://radaruol.uol.com.br/radaruol.html?pg=a>, <http://radix.ibest.com.br/busca/Categoria?pagina=1&catg=arteecultura4.Co&categoria=02>, <http://www.cade.com.br/culhist.html>, <http://www.cade.com.br/cultura.htm>, <http://www.cade.com.br/educacao.htm>, <http://www.cade.com.br/referen.htm>, <http://www.radix.com.br/>, <http://www.radix.com.br/busca/Categoria?pagina=1&catg=cienciaainfo1.Co>, <http://www.radix.com.br/busca/Categoria?pagina=1&catg=cienciasocial1.Co>, <http://www.radix.com.br/busca/Categoria?pagina=1&catg=globalizacao1.Co>, <http://www.radix.com.br/busca/Categoria?pagina=1&catg=historia8.Co>, <http://www.radix.com.br/busca/frequencia>, <http://www.radix.com.br/busca/hits>, <http://www.todobr.com.br/>, and <http://www.uol.com.br/bibliot>. At the bottom of the page, there is a footer with the text 'Internet Archive - Archive-It Web UI 4.9-SNAPSHOT-prod-20140731-2148' and links for 'Help', 'Settings', and 'Submit a Support Question'.

É importante notar que cada website foi arquivado várias vezes ao longo do período, pois o objetivo era o de incluir todas as capturas disponíveis, possibilitando o acompanhamento do que poderia ser alguma lógica contida na programação dos spiders. A frequência e a data de cada varredura estão diretamente relacionadas aos mecanismos e padrões de funcionamento do próprio Internet Archive que, por sua vez, recebe os dados da empresa Alexa, como já foi mencionado.

O projeto propunha inicialmente uma coleção das principais ferramentas de busca do Brasil, entre 1997 e 2005. O marco inicial escolhido, o ano de 1997, corresponde à data das primeiras coletas realizadas pelo Internet Archive. E o marco final, o ano de 2005, coincide com a venda da empresa Todo.br para o Google, o que sinaliza o fechamento de uma fase nesta área das ferramentas de busca no Brasil. A partir daí, o Yahoo! e, principalmente o Google, dominaram totalmente o mercado. Ao longo das interlocuções, o Internet Archive propôs estender este marco cronológico final até o momento da última varredura dos sites de busca selecionados,

realizada em 2013. Inicialmente tive dúvidas, mas decidi aceitar, compreendendo que a função da coleção era também ser útil a outros pesquisadores, cujos projetos de pesquisa poderiam tratar de um período mais recente.

Como já apontamos, a coleção constitui um corpus com ferramentas de busca brasileiras, focando nas temáticas da história e da cultura. Importa notar que apenas duas dentre as ferramentas que compõem o corpus – o Cadê e o Radix – atuavam segundo o modelo de diretórios, funcionando como verdadeiros catálogos online, o que permitiu realizar a coleta de páginas com as temáticas específicas que delimitamos. A etapa de identificação das categorias nativas do período, bem como a coleta de cada link inicial nestas duas ferramentas, foram feitas utilizando a WayBack Machine e de modo absolutamente manual. Sabemos que a pesquisa histórica é tradicionalmente realizada por olhos e mãos humanas. Mas com a crescente escalada das informações em bits, cada vez mais os pesquisadores buscam utilizar ferramentas computacionais que propiciam a varredura de vastos agregados de dados. Porém, no caso da Wayback Machine esta opção não estava disponível⁷².

Há diferenças significativas entre as páginas que constituem os pontos de partida, as páginas semente (*seeds*), de cada uma das ferramentas. A coleção possui *seeds* com profundidades muito variáveis. Por exemplo, o Todo.br não fazia uso das categorias, uma vez que possuía um sistema de busca totalmente automatizado. No caso desta ferramenta, o que coletamos foi então apenas a interface de entrada, não sendo possível emular as buscas tal como realizadas na época. Na entrevista que fizemos com um dos criadores do Todo.br, o prof. Dr. Altigram Soares, ficou claro que não havia uma preocupação com a preservação na época de criação da plataforma. A própria tecnologia empregada não teria como dar conta deste propósito, como afirma Soares: “o *TodoBR* foi desenvolvido para ser um laboratório. (...) Por exemplo, o nosso processador de consultas, ele rodava em um PC, para você ter uma ideia” (SOARES, 2013)⁷³.

Tendo selecionado os endereços das páginas que gostaríamos de inserir na coleção, enviamos a listagem completa para a equipe do Archive-It . Mas, nada é simples no universo das

72 No campo das chamadas Humanidades Digitais, que buscam utilizar os computadores em todas as suas funcionalidades, o debate entre a forma tradicional de pesquisa, chamada em inglês de “close reading”, e a leitura automatizada (distant reading) tem sido um tópico relevante nos debates atuais. Ver, por exemplo, Liu, Alan. “The State of the Digital Humanities: A Report and a Critique.” *Arts and Humanities in Higher Education* 2.1-2 (December 2011): 8-41.

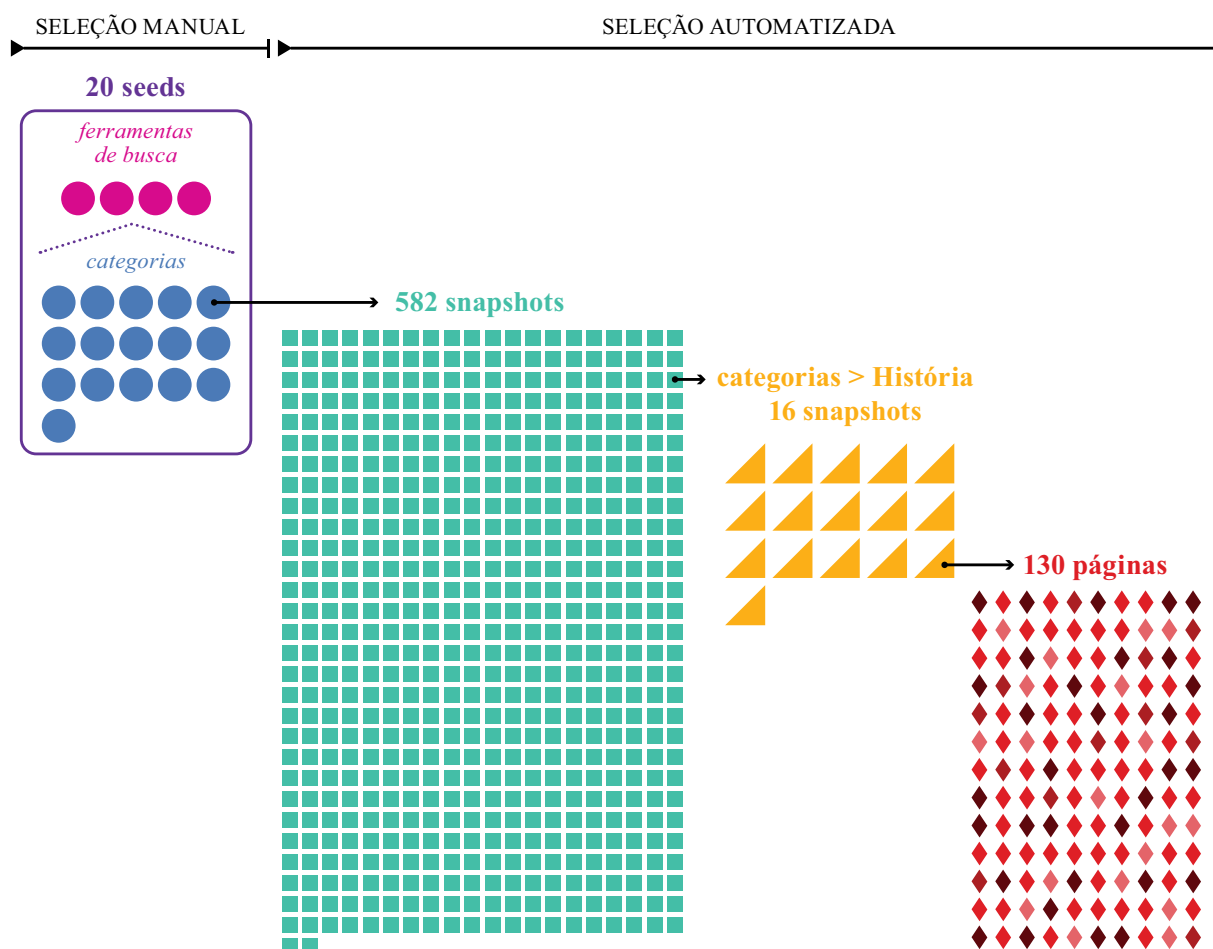
73 Entre 2003 e 2004, no entanto, um banco de dados foi criado para fins experimentais, e talvez ainda esteja sendo utilizado com fins educativos, segundo o entrevistado, na Universidade Federal do Amazonas. Altigram Soares se refere ao pesquisador Edleno Moura, mas não conseguimos entrar em contato para checar a existência do banco de dados. É certamente uma via interessante para futuras pesquisas, que busquem diagnosticar que dados podem ser ainda preservados desta fase inicial da web no Brasil.

páginas arquivadas. O paradigma do Google, de busca e quase mágica aparição de resultados não funciona no universo paralelo da WayBack Machine. O prazo do Archive-It é de no mínimo 60 dias para a coleta de um conjunto de websites. Além disso, o que poderia parecer simples, como a inclusão de mais endereços, não pode ser realizado a curto prazo. Explicando melhor, na medida em que comecei a listar os sites, fui, paralelamente, coletando novas referências, pesquisando mais sobre cada ferramenta selecionada e realizando entrevistas que possibilitaram o acesso a outras informações. Esta investigação indicou a necessidade de incluir outros websites, com o objetivo de compor um cenário mais amplo e, assim, ser capaz de fornecer um contexto da própria coleção. Vimos, por exemplo, a necessidade de incluir sites como o *Aonde* e o *Achei* que, por não estarem hospedados no Brasil, não foram inicialmente agregados à coleção, e sites brasileiros como o *Zeek*, que verificamos ter tido uma importância significativa. De acordo com as últimas comunicações com a equipe do Archive-It, esse novo conjunto de websites só poderá ser adicionado no decorrer do segundo semestre de 2014.

4.2.1 Um percurso na coleção

Neste item pretendemos descrever a coleção através de um percurso que se inicia na página principal da interface do Archive-It. A proposta é tentar dar a ver como os dados foram agregados e, assim, explicar melhor, a arquitetura desta coleção. O anexo digital deste trabalho contém um breve vídeo que busca mostrar um pouco como se dá a navegação entre as páginas passadas da web. A captura dos cliques na plataforma do Archive-It e da WayBack Machine procura apresentar em forma de vídeo o que estamos relatando aqui entre textos e ilustrações. O website escolhido para o vídeo foi o *Cadê*, por nos permitir uma exploração mais ampla de conteúdos e qualidades das páginas arquivadas. A imagem abaixo é uma tentativa de visualização da arquitetura da coleção a partir deste exemplo, procurando mostrar as distintas formas de curadoria (manual ou automatizada) e, ao mesmo tempo, a variação do número de páginas retornadas.

Imagem 32 - Esquema: visualização da arquitetura da coleção na interface do website *Archive-It*



Cada snapshot corresponde ao armazenamento de uma página em determinado ponto no tempo, ou seja, corresponde a uma varredura realizada pelo Internet Archive e transferida via Archive-It para a nossa coleção. É possível se depreender da visualização que a coleção é muito mais rica no primeiro nível e se torna escassa no segundo. A categoria História, por exemplo, foi capturada apenas 16 vezes ao longo dos 16 anos abarcados pela coleção. Mas, o que significa o aumento numérico no próximo nível para 130 páginas? Isto quer dizer que a partir de um snapshot da categoria História da ferramenta de busca Cadê? tivemos acesso a uma lista de 130 páginas.

A maioria das “seeds” contidas no Internet Archive encontram-se relativamente completas, exceto pela ausência de imagens ou banners publicitários. Já no segundo nível da varredura encontramos muitas ruínas digitais em um estado de decomposição maior, ou seja, vários recursos dessas páginas não foram arquivados, deixando as páginas com lacunas ou simplesmente

sem visualização. A imagem abaixo é a tela inicial da categoria História do Cadê?, que diferentemente da página principal só começou a ser arquivada em 2000.

Imagem 33 - Tela inicial da categoria *História do Cadê?* no website *Archive-It*

You are viewing an archived web page, collected at the request of [Universidade Federal do Estado do Rio de Janeiro](#) using [Archive-It](#). This page was captured on 16:54:42 May 08, 1999, and is part of the [Brazilian Web Engines \(1997-2013\)](#). Information on this web page may be out of date. See [All versions](#) of this archived page. [Metadata](#)
[Enable QA](#)

Cadê: Cultura: História

[Opções](#)

Buscar em todo o Cadê? Buscar apenas em Cultura

- [2ª Guerra Mundial](#) - Home page sobre a 2ª Guerra Mundial, com fotos, cronologia da guerra, número aproximado de mortes, sons e filmes.
- [7 Wonders of the World](#) - Página sobre História! As sete maravilhas do mundo, mitologia, grandes civilizações, grandes construções e mais!
- [A Exposição de 1922](#) - Home page sobre a Exposição Internacional de 1922 no Rio de Janeiro, comemorativa do Centenário da Independência. [#07074](#)
- [A História de Canudos](#) - Site sobre a história de Canudos. Cronologia histórica, documentos raros, entrevistas, fotografias, biografias, debates, filmes, livros e pinturas. Biblioteca Pública de Salvador.
- [A Independência do Brasil](#) - Home page sobre a Independência do Brasil. Além disso, contém informações sobre livros e CD-ROMs sobre História, links para criação de Home Pages e links sobre História.
- [Academia: História e Vida Acadêmica](#) - Página de Orlando Ferreira. Informações a respeito de minhas atividades de pesquisa em História da Cultura. Mestrado na Unicamp. História cultural e política inglesa. Literatura gótica.
- [Açúcar da História do Brasil](#) - Coleção de fatos históricos importantes ocorridos no Brasil. Os 100 anos de Canudos. Conheça a Bahia. Salvador, BA.
- [Aço Brasileiro](#) - História do pioneirismo do aço brasileiro, CSN - Companhia Siderúrgica Nacional e Volta Redonda.
- [Aldus Homepage](#) - Página dedicada à obra de J.R.R. Tolkien e aos interessados em História Antiga e Medieval.
- [Ancient Egypt's Home-Page](#) - Um site brasileiro sobre Egíptologia com informações sobre história, arte, cultura, religião e arqueologia. Produzida por Marcio Luiz Rames D'Albuquerque.
- [Anos 50 Home Page](#) - Uma viagem para algum lugar entre 1954 e 1964, a era de ouro onde a juventude de cabelos em desalinho, vestidos rodados, roupas coloridas, jaquetas de couro, rebeldes e os primórdios do Rock n' Roll tomavam conta de sua época.
- [Anos 60 Internet Images](#) - Fotos que marcaram a década de 60 na política, artes, música e na sociedade. Ditadura, primeiro homem na lua, Woodstock, Martin Luther King, John Lennon, Beatles, Kennedy, Che Guevara, Tropicália, Jovem Guarda, cinema nacional e mais.
- [Anti-Exploração Home Page](#) - História geral. Site Marxista. Críticas, tópicos como Segunda Guerra Mundial, socialismo, capitalismo, neoliberalismo, nossos presidentes, Collor de Mello e mais.
- [Arquitetura](#) - Site desenvolvido para pessoas interessadas em Arqueologia e História em geral. [#07074](#)
- [Arquivo de História Social Edgar Rodrigues](#) - Reúne informações sobre a história dos movimentos libertários no Brasil e no mundo, possuindo também informações sobre a obra do pesquisador de história social Edgar Rodrigues.
- [Arquivo Público do Estado do Pará](#) - Arquivo histórico com documentação do século XVII, XVIII, XIX e XX. Coleção de documentos de importância para a história da Amazônia Colonial e Imperial brasileira.
- [Arthur Franco: A Idade das Luzes e os Megalitos de Hy-Brasil](#) - A descoberta do sítio megalítico do Sul do Brasil e as lendas sobre o Reino de Hy-Brasil pelos Gregos, Fenícios, Frígios, Cários, Hebreus e Celtas. Conheça a História por trás da História, na Filosofia Religiosa...
- [As Aventuras de Avelar, o Malhado](#) - Ficção com fundo histórico. Lutas pela independência da América Espanhola no primeiro quartel do século XIX. Referências a Venezuela, Ibaneiros, Simón Bolívar e José Antônio Páez. Links acadêmicos e outros.
- [Bandeiras Históricas Luso-Brasileiras](#) - Página sobre as bandeiras utilizadas pelo Brasil durante os períodos colonial (1500-1822), imperial (1822-1889) e republicano (1889-...). Links.
- [Biografias](#) - Histórias de personagens que fizeram história. [#07074](#)
- [Ceará: O Processo Civilizador](#) - A descoberta do Brasil no Ceará pelo espanhol Vicente Pinzon antes de Cabral a história da formação e colonização do estado do Ceará. Fotografia da maquete de Fortaleza em 1726.
- [Cens Medievais](#) - Banco de imagens sobre o período da Idade Média (medieval). História, feudalismo e Idade Média.
- [Centro Nacional de Pesquisas Históricas](#) - Holocausto Judeu, Segunda Guerra, Nazismo, Revisionismo, links e livros revisionistas.
- [Civilizações Desaparecidas Web Page](#) - Página voltada aos enigmas e curiosidades das grandes e velhas civilizações desaparecidas. Atlântida, Lemúria etc...
- [Clio - Assessoria em Pesquisa Histórica da Bahia](#) - História, fontes, arquivos, mapas, bibliotecas, dissertações, Colônia, Império, Canudos, república, escravidão etc... Salvador, BA. [#07074](#)
- [Coca-Cola Anúncios Históricos](#) - Apenas anúncios impressos da Coca-Cola, desde o primeiro brasileiro em 1944. Em português, espanhol, inglês e alemão.
- [Conflitos Árabes - Israelenses](#) - Página sobre os conflitos entre Israel e os Estados Árabes. Biografia dos líderes, história das guerras, o que é o OLP, fundamentalismo, terrorismo, movimento nacional palestino, sionismo etc...

Fonte: Archive-It. Brazilian Web Engines

Cada endereço listado foi incluído na página manualmente, pois na época, como detalharemos mais adiante, o Cadê? seguia o modelo de diretório onde cada página adicionada era feita através de uma operação manual de inclusão. Hoje, boa parte destes endereços não pode ser acessada, uma vez que não foi alvo da seleção inicial do Internet Archive. Um exemplo é a página Anti-Exploração da qual só restou esta ruína:

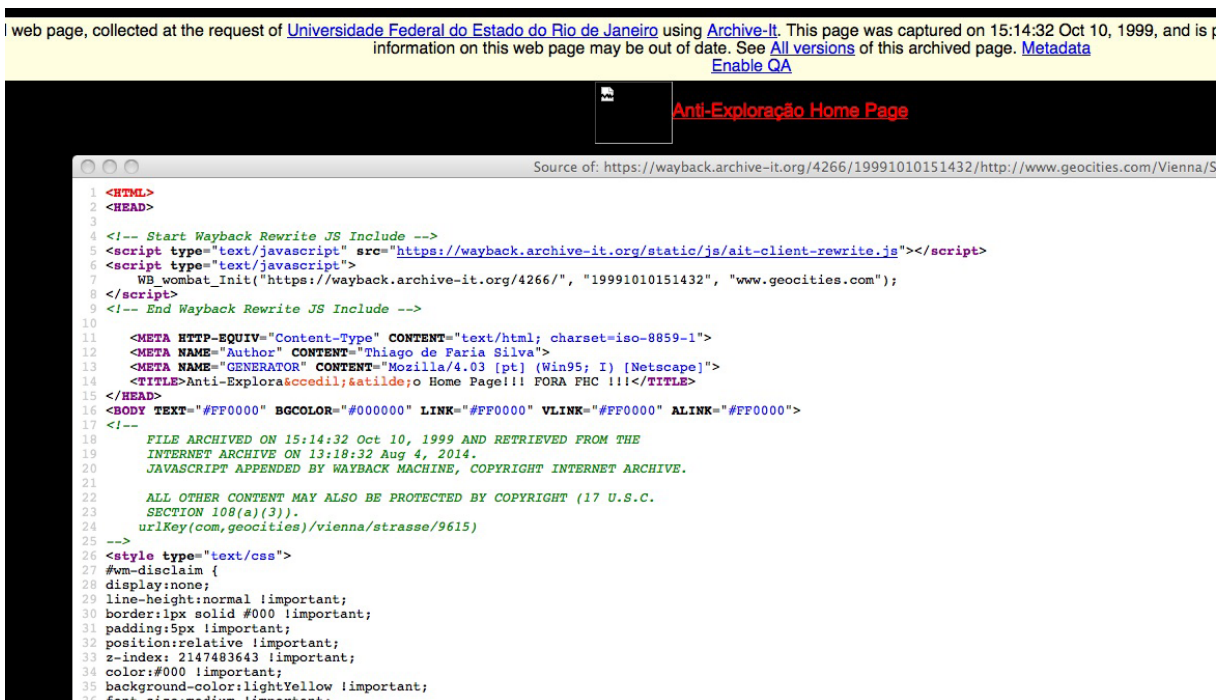
Imagem 34 - Ruína da página *Anti-Explocação* no website *Internet Archive*



Fonte: Internet Archive.

A partir desta imagem pouco se pode depreender deste website, mas se abrirmos o código da página em HTML (HyperText Markup Language) podemos descobrir ao menos duas informações importantes, o autor da página (Tiago de Faria Silva) e a descrição da página escrita pelo autor: “Fora FHC”.

Imagem 35 - Informações do código da página *Anti-Explocação* em HTML no website *Internet Archive*



The image shows a screenshot of a web browser displaying the source code of a page. The browser's address bar shows the URL: `Source of: https://wayback.archive-it.org/4266/19991010151432/http://www.geocities.com/Vienna/`. The page title is "Anti-Explocação Home Page". The source code is displayed in a monospaced font, showing HTML and JavaScript tags. The code includes a meta tag for the author (Thiago de Faria Silva), a meta tag for the generator (Mozilla/4.03 [pt] (Win95; I) [Netscape]), and a title tag: "Anti-Explocação Home Page!!! FORA FHC !!!". There are also several comments and style tags, including a disclaimer and a font size setting.

```

1 <HTML>
2 <HEAD>
3
4 <!-- Start Wayback Rewrite JS Include -->
5 <script type="text/javascript" src="https://wayback.archive-it.org/static/js/ait-client-rewrite.js"></script>
6 <script type="text/javascript">
7   WB_wombat_Init("https://wayback.archive-it.org/4266/", "19991010151432", "www.geocities.com");
8 </script>
9 <!-- End Wayback Rewrite JS Include -->
10
11 <META HTTP-EQUIV="Content-Type" CONTENT="text/html; charset=iso-8859-1">
12 <META NAME="Author" CONTENT="Thiago de Faria Silva">
13 <META NAME="GENERATOR" CONTENT="Mozilla/4.03 [pt] (Win95; I) [Netscape]">
14 <TITLE>Anti-Explocação Home Page!!! FORA FHC !!!</TITLE>
15 </HEAD>
16 <BODY TEXT="#FF0000" BGCOLOR="#000000" LINK="#FF0000" VLINK="#FF0000" ALINK="#FF0000">
17 <!--
18   FILE ARCHIVED ON 15:14:32 Oct 10, 1999 AND RETRIEVED FROM THE
19   INTERNET ARCHIVE ON 13:18:32 Aug 4, 2014.
20   JAVASCRIPT APPENDED BY WAYBACK MACHINE, COPYRIGHT INTERNET ARCHIVE.
21
22   ALL OTHER CONTENT MAY ALSO BE PROTECTED BY COPYRIGHT (17 U.S.C.
23   SECTION 108(a)(3)).
24   urlKey(com,geocities)/vienna/strasse/9615)
25 -->
26 <style type="text/css">
27 #wm-disclaim {
28 display:none;
29 line-height:normal !important;
30 border:1px solid #000 !important;
31 padding:5px !important;
32 position:relative !important;
33 z-index: 2147483643 !important;
34 color:#000 !important;
35 background-color:lightYellow !important;
36 font-size:medium !important;

```

Fonte: Internet Archive

Percorrer o *Acervo digital das principais ferramentas de busca brasileiras para História e Cultura (1997 a 2003)*, assim como outras coleções digitais criadas com o apoio com o Internet Archive, é, portanto, deparar-se com informações codificadas. Mas, ao mesmo tempo, a plataforma do Archive-It possibilita o acesso aos dados através da busca por palavras. Desse modo, a coleção disponibiliza uma série de informações que podem ser utilizadas por qualquer pesquisador interessado nos fragmentos da história da internet no Brasil, em especial na atuação das ferramentas de busca antes da chamada Era Google. Ao descrevermos um percurso na coleção, procuramos dar a ver sua arquitetura, mas é preciso chamar atenção para o fato de que apresentamos apenas uma entre tantas outras possíveis navegações.

É até possível imaginar um sistema de emulação das interfaces tecnológicas do passado recente da internet onde fosse possível também supor os resultados de determinadas buscas, digamos, de uma década atrás. Este, porém, seria um projeto de fôlego, envolvendo todo um aparato tecnológico e equipe especializada. A nossa massa de dados sobre as histórias das ferramentas de busca é bem mais modesta. A intenção foi criar uma coleção que possibilitasse uma investigação das ferramentas de busca brasileiras em seus primeiros desenvolvimentos, focando especialmente no modo como estas ferramentas davam acesso a conteúdos relacionados à história e à cultura.

A criação desta coleção significou mergulhar numa investigação sobre a materialidade dos registros em código binário. A aposta é de que os instantâneos dos primórdios da internet brasileira são vestígios de um sistema de classificação e podem revelar processos históricos por vezes esquecidos diante da proeminência das telas de buscas contemporâneas. Ao navegar pela coleção é possível perceber, por exemplo, um processo de desaparecimento dos diretórios e das classificações hierárquicas organizadas por pessoas, crescentemente substituídas pelas páginas com um visual clean - a popular caixa de buscas do Google. A coleção apresenta-se, assim, em certa medida, como um gesto no sentido de desnaturalizar as telas do presente, chamando atenção para a historicidade das tecnologias a nossa volta. Este gesto segue a trilha de alguns pesquisadores como Richard Rogers, coordenador do Digital Methods Initiative, que chama atenção para a necessidade de se estudar os objetos nascidos digitais a partir de sua dinâmica própria, dando outros usos às ferramentas e explicitando o processo histórico (2013:25).

Por outro lado, ao longo do percurso percebemos a necessidade de complementar a documentação sobre as ferramentas de busca brasileiras com a investigação da memória oral, por meio de entrevistas com atores relevantes no período. O próximo item tratará da temática da coleção, utilizando como evidência os relatos, e propondo, desta maneira, uma metodologia que abarque as fontes nascidas digitais em convergência com a memória oral, não apenas para suprimir as lacunas, mas, sobretudo, para lançar novos olhares sobre as páginas arquivadas em dígitos.

4.2.2 O caso de um catálogo brasileiro da web 1.0: o portal de buscas Cadê?

Como a coleção se constituiu a partir das varreduras realizadas pelo Internet Archive, conforme descrito nos itens anteriores, não é possível recuperar com profundidade as várias camadas de links. Além disso, as lacunas são bastante significativas: faltam imagens, banners, links... Está claro que ainda assim o pesquisador poderia optar por fazer uma investigação apenas através das fontes digitais e trabalhar as lacunas documentais a partir de outras perspectivas, pois sempre haverá lacunas em qualquer documentação. A opção de ampliar a coleta de fontes, incluindo reportagens de jornal, numa primeira fase, e, depois, entrevistas com os principais responsáveis pelas ferramentas de busca, não teve como intuito, portanto, construir uma narrativa que desse conta de uma suposta totalidade.

Ao perguntar acerca da própria lógica das ferramentas, vimos nas fontes orais a possibilidade de sondar como estas construções estavam associadas a práticas correntes do final do século XX no Brasil. Por exemplo, uma das perguntas que a documentação disponível não

conseguia responder diz respeito ao funcionamento do retorno das pesquisas. Isto é: quais mecanismos ou filtros cada ferramenta utilizava para ordenar os resultados finais oferecidos aos usuários. E ainda: como a publicidade era inserida? Havia rankings de popularidade? Todas estas perguntas puderam ser abordadas nas entrevistas, que foram fundamentais para entender melhor o contexto brasileiro da constituição das diversas ferramentas.

Nesta parte do trabalho iremos inicialmente abordar os diferentes modelos de busca existentes no período, o diretório e o engenho/máquina, relacionando os casos escolhidos. Em seguida, com o objetivo de aprofundar em um dos casos, nos deteremos na experiência do Cadê?, escolhido em razão da sua popularidade entre os anos de 1995 e 2001.

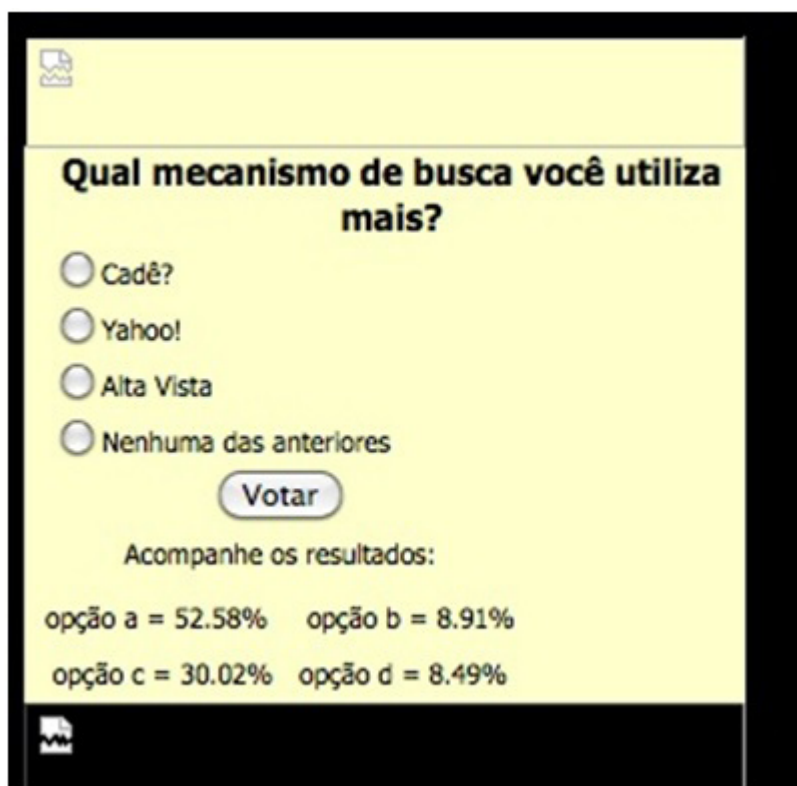
4.2.3 O Cadê? a partir do relato de um dos seus fundadores

Várias reportagens entre os anos 1997 e 1998 apontam o Cadê? como a principal ferramenta de buscas utilizada no Brasil ou, nos termos da época, “o queridinho dos internautas”⁷⁴. De acordo com Gustavo Viberti, um dos fundadores da empresa, naquele período o maior concorrente do Cadê? era o Universo Online (UOL) da Folha de São Paulo, o maior portal de conteúdos e que também atuava como provedor de acesso à rede. Em maio de 1999⁷⁵, uma enquete realizada pela revista Internet.Br apontava o Cadê? como a ferramenta de busca mais utilizada pelos leitores do periódico. As entrevistas realizadas para essa pesquisa confirmam a importância do Cadê? no cenário da internet brasileira, embora não possamos afirmar categoricamente se esta ferramenta era de fato a primeira da lista.

⁷⁴ VASCONCELOS, Nelson. *Tem alguém perdido aí?* Informática&Etc. Jornal O Globo. 14/04/1997.

⁷⁵ A revista Internet.Br publicou uma enquete em 8/05/1999, onde a ferramenta de busca mais votada era o Cadê?. Disponível no Internet Archive: <http://web.archive.org/web/19981212025957/http://www.internetbr.com.br/>

Imagem 36 - Enquete realizada pela revista *Internet.Br* apontava o *Cadê?* como a ferramenta de busca mais utilizadas pelos leitores do periódico



Qual mecanismo de busca você utiliza mais?

Cadê?

Yahoo!

Alta Vista

Nenhuma das anteriores

Votar

Acompanhe os resultados:

opção a = 52.58% opção b = 8.91%

opção c = 30.02% opção d = 8.49%

Fonte: Internet Archive.

Para entender um pouco o sucesso alcançado pelo Cadê? é preciso pesquisar o contexto da época que estava assistindo ao nascimento da web, e também o que buscavam os fundadores de uma das primeiras ferramentas de busca brasileiras.

A Rede Nacional de Pesquisa (RNP) exerceu um papel decisivo na coordenação de esforços de infra-estrutura da rede no Brasil, mas também mantinha uma presença importante na própria web⁷⁶. O website da RNP contou com o Yaih, a versão brasileira do Yahoo, que foi uma das fontes para a criação de outros diretórios do Brasil como por exemplo o Cadê. Abaixo pode-se ver algumas imagens da web brasileira em 1997:

⁷⁶ Ver, Carvalho, Marcelo Sávio Revoredo Menezes de. *A trajetória da Internet no Brasil: do surgimento das redes de computadores à instituição dos mecanismos de governança*. Rio de Janeiro, Dissertação de mestrado Universidade Federal do Rio de Janeiro, COPPE, 2006, p.84.

Imagem 37 - Algumas imagens da web brasileira em 1997 nos arquivos do website *Internet Archive*



Ciências Humanas

• Antropologia <small>NEM - (1 URL)</small>	• Listas de discussão@ <small>NEM - (5 URL)</small>
• Antroposofia <small>NEM - (1 URL)</small>	• Literatura <small>NEM - (3 DIR) - (42 URL)</small>
• Direito <small>NEM - (2 DIR) - (53 URL)</small>	• Logosofia <small>NEM - (1 URL)</small>
• Economia <small>NEM - (2 URL)</small>	• Política <small>NEM - (1 DIR) - (5 URL)</small>
• Eventos <small>NEM - (1 URL)</small>	• Psicologia <small>NEM - (6 DIR) - (22 URL)</small>
• Filosofia <small>NEM - (14 URL)</small>	• Revistas@ <small>NEM - (1 DIR) - (3 URL)</small>
• Geografia <small>NEM - (2 DIR) - (3 URL)</small>	• Serviço Social <small>NEM - (1 URL)</small>
• História <small>NEM - (1 DIR) - (15 URL)</small>	• Sociologia <small>NEM - (1 DIR) - (6 URL)</small>
• Ideologias <small>NEM - (2 DIR) - (1 URL)</small>	• Teologia <small>NEM - (1 URL)</small>
• Linguística <small>NEM - (7 URL)</small>	

- [Auto Domínio](#) - Página de auto domínio. Auto conhecimento e ajuda. Pensamento Construtivo e Proativo.
- [Biblioteca Virtual de Estudos Culturais \(CNPq/UFRR/PACC\)](#) - Site que reúne informação de diversas naturezas sobre Estudos Culturais no Brasil e no exterior: cursos de pós-graduação, periódicos, instituições de pesquisa, etc.
- [Companhia Vale Do Rio Doce](#) - Abaixo Assinado contra a venda da CVRD e um breve texto sobre essa importante empresa nacional
- [Monastério da Ordem dos Místicos](#) - Aprenda sobre Mantra, Espiritualidade, Cosmos, Energia, Parapsicologia, Psicologia, Radiestesia, terapia natural, Cibernética, Aurea, Metapsíquica, Bioenergia, Karma, Dharma, Universo, Quântica.
- [O Mundo do Esperanto](#) - Aqui você encontra toda informação sobre a Língua Internacional; sua criação e desenvolvimento; a importância de uma língua neutra no mundo de hoje; aprendizagem, etc.
- [PNMAPC - Simpósio: Política Nacional de Meio Ambiente e Patrimônio Cultural](#) - Repercussões dos dez anos da Resolução Conama n.001/86 sobre a pesquisa e a gestão dos recursos culturais no Brasil.

[Centro de Informações Internet Brasil](#)
[RNP - Rede Nacional de Pesquisa](#)
webmaster@ci.rnp.br

Fonte: Internet Archive

Ainda no último ano da década de 1980 o Instituto Brasileiro de Análises Sociais e Econômicas (Ibase) criou o primeiro serviço de internet do Brasil com conexões entre as Organizações Não-Governamentais e serviços de e-mail. Em 1989, quarenta entidades participavam da rede do Ibase, a Alternex⁷⁷. Em 1991, foi criada uma rede acadêmica entre Rio e São Paulo através da FAPESP, fundação que ficou responsável pelo registro de domínios no Brasil até 2005⁷⁸. Embora distante dos principais centros de desenvolvimento tecnológico a trajetória da expansão tecnológica no país foi também espetacular. Em 1998, apesar do avanço nos principais centros urbanos, o Brasil ainda estava bastante distante do desenvolvimento da rede se comparado aos países de maior produto interno bruto do mundo, ocupando o 18º lugar no número de *hosts(servidores)* Em menos de 10 anos o número de *hosts* saltou de 117.200 para mais de oito milhões em julho de 2007, passando a ocupar a nona posição no ranking mundial (na frente de países como Canadá e Inglaterra) e a primeira posição na América do Sul⁷⁹.

⁷⁷ Sobre o pioneirismo do Ibase, ver o site da instituição: <http://www.ibase.org.br/modules.php?name=Conteudo&pid=1210> (acessado em agosto de 2007)

⁷⁸ Até 2005 a FAPESP foi a responsável pelo registro de domínios e IPs(protocolos da internet) no Brasil. Em dezembro daquele ano, devido ao aumento exponencial do volume de registros, esta responsabilidade foi transferida para o Núcleo de Informação e Coordenação do Ponto BR (<http://www.nic.br/sobre-nic/index.htm>). Esta transferência foi fruto de um longo processo já em andamento compartilhado com o Comitê Gestor da Internet no Brasil.

⁷⁹ Dados do Centro de Estudos do Comitê Gestor da Internet no Brasil. Disponível em: <http://www.cetic.br/index.html> (acessado10/10/2013).

A inspiração do Cadê? foi também o Yahoo!. As imagens abaixo explicitam as semelhanças entre os diretórios do Yahoo! e do Cadê?, no ano de 1997.

Imagem 38 - Página inicial do Yahoo! em 1997

YAHOO!

New Cool Today's News More Yahoos

Wimbledon SURPLUS DIRECT Win a 3000GT VR-4 Click Me Summer Movie Preview

Search options

Yellow Pages - People Search - Maps - Classifieds - News - Stock Quotes - Sports Scores

- **Arts and Humanities**
[Architecture](#), [Photography](#), [Literature...](#)
- **Business and Economy [Xtra!]**
[Companies](#), [Investing](#), [Employment...](#)
- **Computers and Internet [Xtra!]**
[Internet](#), [WWW](#), [Software](#), [Multimedia...](#)
- **Education**
[Universities](#), [K-12](#), [College Entrance...](#)
- **Entertainment [Xtra!]**
[Cool Links](#), [Movies](#), [Music](#), [Humor...](#)
- **Government**
[Military](#), [Politics \[Xtra!\]](#), [Law](#), [Taxes...](#)
- **Health [Xtra!]**
[Medicine](#), [Drugs](#), [Diseases](#), [Fitness...](#)
- **News and Media [Xtra!]**
[Current Events](#), [Magazines](#), [TV](#), [Newspapers...](#)
- **Recreation and Sports [Xtra!]**
[Sports](#), [Games](#), [Travel](#), [Autos](#), [Outdoors...](#)
- **Reference**
[Libraries](#), [Dictionaries](#), [Phone Numbers...](#)
- **Regional**
[Countries](#), [Regions](#), [U.S. States...](#)
- **Science**
[CS](#), [Biology](#), [Astronomy](#), [Engineering...](#)
- **Social Science**
[Anthropology](#), [Sociology](#), [Economics...](#)
- **Society and Culture**
[People](#), [Environment](#), [Religion...](#)

Imagem 39 - Página inicial do Cadê? em 1997

novos **inclusões** **cadê?** **eventos** **veja** **?**

Consulta Busca [Informações](#)

Ciência e Tecnologia
[Institutos](#), [Centros de Pesquisa](#)

Cultura
[Museus](#), [Música](#), [Personalidades](#)

Esportes
[Automobilismo](#), [Futebol](#)

Governo
[Estados](#), [Prefeituras](#)

Informática
[BBS](#), [Empresas](#), [Software](#)

Lazer
[Bares e Restaurantes](#), [Turismo](#), [TV](#)

Referência
[Bibliotecas](#), [Dicionários](#)

Serviços
[Propaganda](#), [Tradutores](#)

Compras
[CDs](#), [Informática](#), [Livros](#)

Educação
[Cursos](#), [Escolas](#), [Universidades](#)

Finanças
[Bancos](#), [Bolsas](#), [Seguros](#)

Indústria e Comércio
[Automóveis](#), [Telecomunicações](#)

Internet
[Provedores \(DF, MG, RJ, SP, ...\)](#)

Notícias
[Classificados](#), [Jornais](#), [Revistas](#)

Saúde
[Hospitais](#), [Instituições](#)

Sociedade
[Meio Ambiente](#), [ONGs](#), [Pessoais](#)

[Cadê? Remoto](#) | [Mapa](#) | [Download](#) | [Ao Vivo](#) | [Estatísticas](#) | [E-mail](#) | [Propaganda no Cadê?](#)

Fonte: Internet Archive

O Yahoo! (acrônimo para Yet Another Hierarchical Officious Oracle) foi criado em 1994 por dois engenheiros e pós-graduandos da Stanford University: Jerry Yang and David Filo. Eles começaram a empresa a partir de um website pessoal que muito rapidamente se transformou em um dos principais portais de acesso à rede mundial de computadores⁸⁰. O modelo da busca em um diretório hierárquico com categorias de caráter popular tornou-se um modelo fundamental de buscas na web. A folksonomia, padrão/modelo de organização da informação adotado pelo

⁸⁰ Sobre o sucesso do Yahoo! Ver ROGERS (2010).

Yahoo!, é uma forma de arranjar as categorias que se diferencia da classificação tradicional das bibliotecas, realizada de acordo com o sistema CDD (Classificação Decimal de Dewey). Há uma maior flexibilidade nas categorias adotadas, ao mesmo tempo em que a adição de sub-categorias é bastante frequente. O Yahoo! Logo passou a receber inscrições de websites para o seu diretório e ampliou sua política de propaganda. Foi um dos primeiros empreendimentos de tecnologia desenvolvidos por estudantes que alcançou sucesso comercial. Com apenas um ano de funcionamento, a página foi acessada um milhão de vezes (FRAGOSO, 2007:5).

Ao relatar o contexto profissional e pessoal em meio ao qual tomou forma a ideia do Cadê?, Gustavo Viberti deixa clara a inspiração no Yahoo!. Viberti conta que era um engenheiro empregado em uma grande empresa à procura de novos rumos profissionais. Decidiu tirar férias para pensar e redefinir a carreira. E foi justamente durante este período que ele começou a usar mais a internet, percebendo, surpreso, que ela já havia extrapolado o meio acadêmico e se abria como um universo a ser explorado: “(...) alguma coisa seria necessária para que as pessoas soubessem o que era aquilo – o que era a internet. O Yahoo!, que era um exemplo lá fora, estava começando também.” (VIBERTI, 2014)⁸¹. Gustavo lançou então, ainda durante seus dias de férias - em outubro de 1995 -, o Cadê?, site de buscas com, inicialmente, 300 páginas brasileiras e tecnologia 100% nacional. Ele conta que se pautou bastante no Yahoo! para montar o diretório de categorias e que rapidamente as pessoas começaram a entrar no site e enviar páginas novas para inclusão: “No início eram 10 páginas, 20 ou 30 por dia. Mas, em pouco tempo, em dezembro, eu já recebia por dia 150 páginas” (VIBERTI, 2014)⁸².

O rápido crescimento do Cadê? fez Viberti convidar Fábio Oliveira para entrar como sócio. Logo contrataram um editor que recebia as sugestões de páginas, pesquisava, fazia uma resenha sobre cada uma delas e, finalmente, as incluía no site. Em outubro de 1995, eram 300 endereços cadastrados no site, mas já no início de 1996 chegaram a 1200. Com a rápida visibilidade alcançada pela empresa, três meses depois do início do Cadê? apareceram, para surpresa dos sócios, os primeiros interessados em investir no negócio. “Não se imaginava uma coisa assim no Brasil, naquele momento. A internet não era importante, nem lá fora era muito importante! Os IPOs não tinham começado, então era uma coisa um pouco fora, assim, da expectativa nossa conseguir um investidor em tão pouco tempo”, conta Viberti (2014). Outro efeito do rápido crescimento foi a impossibilidade do provedor usado suportar o volume do tráfego em

81 Entrevista, Anexo I.

82 A página do Cadê? neste primeiro ano não foi arquivada pelo Internet Archive que, como mencionamos, começou a receber os dados da Alexa somente no final de 1996. Na entrevista, Gustavo Viberti declarou não possuir a documentação digital deste período.

expansão do portal Cadê?. A chegada dos investidores permitiu que o problema fosse superado em uma época na qual o custo de acesso era muito caro.

O mercado também passou a reconhecer no Cadê? um meio eficiente para se chegar ao público e já no início de 1996 o site fechou dois grandes e importantes contratos: um com o Ministério da Educação para divulgar o Provão, espécie de Enem da época, e outro com a Amazon. Viberti conta que o primeiro contrato foi importante para entender qual era o alcance do Cadê?. Isto porque o Ministério da Educação usou apenas o site para divulgar informações, através de um banner publicitário, e convidar as pessoas para um *chat* que seria feito com o ministro. No dia e hora combinados, o número de jovens que entrou no site superou em muito as expectativas, o que fez com que o *chat* caísse todo o tempo. O episódio mostrou, assim, o grau de difusão do Cadê?. A Amazon, por sua vez, escolheu o site para fazer a primeira campanha de venda de livros fora dos Estados Unidos.

O crescimento do Cadê? está ligado a um contexto mais amplo de expansão das ferramentas de busca na internet. Os usuários passaram a usar esses buscadores de maneira gratuita, já que a forma de financiamento foi vinculada à publicidade. O estudo empírico realizado por Dain e Petersen, já em um momento posterior, demonstrou que em apenas uma semana de agosto de 2005 as ferramentas de busca exibiram 13 bilhões de resultados patrocinados.

O processo de catalogação do Cadê? era totalmente manual, ou seja, cada endereço recebido era checado por um funcionário do site e só então ingressava no banco de dados. Ao navegar na listagem do Cadê disponível no Internet Archive, me perguntava sobre as formas de filtros na inclusão destes sites. Haveria algum tipo de censura política? Ou algum critério relativo aos gostos e interesses dos criadores? Apenas dois critérios de exclusão eram aplicados: violência explícita e pedofilia. Neste último caso, se eles recebessem uma página com este tipo de conteúdo, denunciavam à Polícia Federal, como explica Viberti no seguinte trecho:

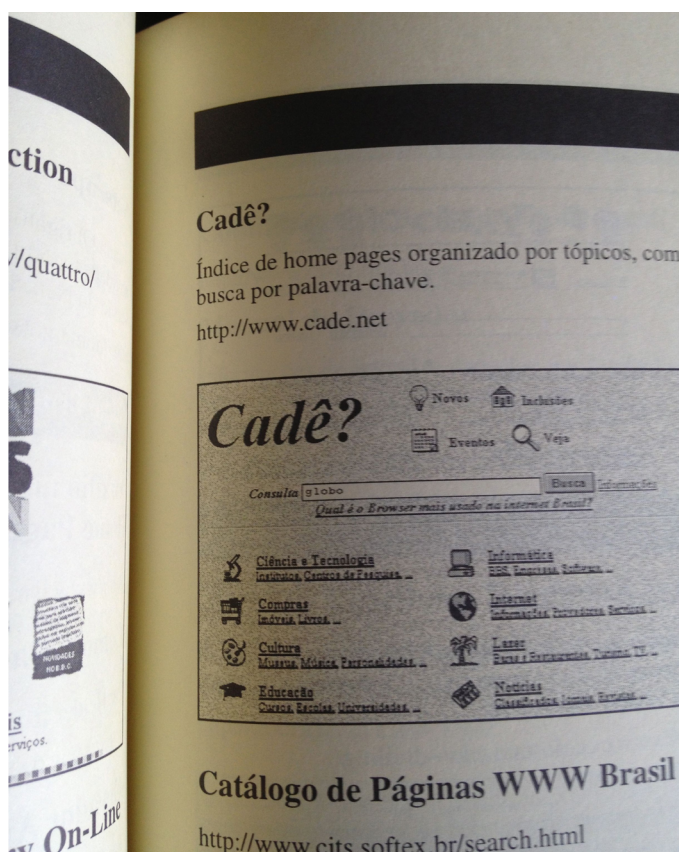
“Quando uma pessoa submetia uma página, os editores entravam na página para dar uma olhada e fazer uma resenha dela. Era um trabalho que demandava algum tempo, mas que evitava que entrassem no Cadê? propagandas e coisas assim. Então, toda resenha era feita pelos editores. Quando eles entravam na página, viam se realmente se tratava daquilo proposto. Se houvesse pornografia infantil, a gente denunciava a página. Tínhamos um canal com a Polícia Federal, um grupo da PF que trata pedofilia, e a gente enviava para eles. E não incluía, claro. Outra era ‘terrorismo’. Uma vez apareceu uma página que ensinava a fazer bombas. Discutíamos sobre liberdade na internet, conteúdo livre, ideias livres, mas quando chegou uma página deste tipo, que ensinava como fazer uma bomba, não dava para incluir. Então, a gente procurava ter

critérios mínimos para, por um lado, não coibir a liberdade daquilo que fosse criado, mas, por outro lado, coisas como pedofilia, ensinar a fazer bomba, não entravam”. (VIBERTI, 2014)

As páginas novas eram incluídas em uma das 16 categorias principais e, dentro destas, entravam em ordem alfabética, tal qual um catálogo. Na medida em que estas páginas incluídas se referissem a um mesmo assunto, novas subcategorias eram formadas e assim por diante. Ao fazerem suas buscas, os internautas recebiam as informações nesta mesma ordem de inclusão. Isto porque o Cadê? não promovia um ranking das páginas mais acessadas, como vieram a fazer outros mecanismos de busca como o Altavista, o Lycos e o Google, que estabeleceram o modelo hoje vigente.

Além das informações contidas no Internet Archive, existem algumas outras publicações que contam um pouco da história do Cadê?. Isso porque Gustavo Viberti não se contentou apenas em usar o meio digital para se comunicar e se valeu também do meio impresso para lançar livros e catálogos. A imagem abaixo traz a referência ao Cadê, no catálogo de 1997.

Imagem 40 - Catálogo de 1997 traz referência ao *Cadê?*



Nos arquivos da Biblioteca Nacional há um catálogo contendo uma listagem de tudo o que havia no Cadê? em 1996/97⁸³. Como o site era muito copiado e, na época, não havia nada na legislação brasileira que regulasse o assunto, a solução jurídica encontrada foi o registro no órgão como forma de proteção⁸⁴.

Mas não foi apenas em busca de proteção legal que o Cadê? passou a publicar o seu conteúdo. Catálogos nos moldes das “listas amarelas” foram editados com o propósito de divulgação e de consulta. O que para o momento atual soa absolutamente estranho (Quem pensaria em divulgar um site através de um catálogo? Quem pensaria em encontrar um site consultando um catálogo?) para aquela época fazia total sentido. A internet ainda era um território em formação e o cruzamento com a mídia impressa constituía possivelmente uma estratégia para consolidá-la. Por outro lado, é interessante pensar também em que medida sistema de catalogação hierárquico por assuntos, que se desenvolve em sites como o Yahoo! e o Cadê?, reverberam sobre os modelos usuais de catalogação e sobre todo um modo de organização da informação ligados à mídia escrita. Em outros termos, significa perguntar em que medida as especificidades do meio digital atingem o meio impresso. Estas e outras questões só puderam aparecer ao nos debruçarmos sobre a documentação escrita, cruzando-a com as fontes nascidas digitais. Antes de enveredarmos em uma reflexão sobre as formas de organizar as informações digitais que constituiriam a base da memória documental aqui pesquisada, iremos prosseguir com a narrativa da construção da empresa Cadê?.

Segundo Viberti, o Cadê?, chegou a ser o site mais acessado no Brasil entre 1997 e 1998 quando cerca de 90% das buscas brasileiras eram feitas por meio dele, porcentagem que equivale às buscas feitas hoje pelo Google. A medição do tráfego na internet é até hoje um campo problemático. Na primeira pesquisa de ranking nacional dos sites, o Cadê? ficou em terceiro lugar (Altavista.com era o primeiro, seguido pelo Bol.com.br) na medição realizada pela Media-metrix.⁸⁵ Mas, apesar do sucesso, os sócios temiam pela saúde financeira do negócio por conta da chegada de grandes empresas ao Brasil e, por esta razão, venderam o Cadê? em 1999 para a Starmedia, uma grande empresa americana. Algum tempo depois o Cadê? acabaria mudando novamente de mãos, sendo vendido ao grupo Yahoo!. Viberti explica:

83 O registro se encontra no Escritório de Direitos Autorais da Biblioteca Nacional e está disponível em: http://www.bn.br/portal/?nu_pagina=25 (acessado em 10/06/2014)

84 Dessa época até o momento atual, o contexto mudou. O País já caminhou bastante na questão legislativa com a sanção da Lei 12.965, de 2014 (chamada Marco Civil), que trata dos princípios, garantias, direitos e deveres para internautas e provedores na rede mundial de computadores no Brasil. O Marco Civil ainda será regulamentado e as consequências certamente serão igualmente importantes para o campo de estudos da memória social.

85 Fonte: Canadian Corporate Newswire.14 de Setembro, 2000. “First-Ever Brazilian Internet & Digital Media Measurement Results Revealed By Media Metrix”. Via Lexis Nexis. Viberti faz referência a uma pesquisa IBOPE de 1997 que não tive acesso direto.

“Naquele momento decidimos vender a empresa porque havia muitas ameaças externas ao nosso capital. A empresa tinha crescido, mas não era tão grande assim para fazer frente a gente muito grande que estava chegando ao Brasil naquele momento. Primeiro começamos a procurar um investidor, alguém que entrasse com algum dinheiro para fazer a empresa crescer mais rápido, mas o pessoal não queria isso, eles queriam realmente comprar tudo. Então eles compraram a empresa em 1999 e nós permanecemos ainda à frente dela até 2000, durante um ano depois da compra, por uma condição contratual que os compradores colocaram” (VIBERTI, 2014).

É interessante perceber que o discurso de Altigram Soares, que participou ativamente na criação do TodoBr, uma ferramenta de caráter mais acadêmico, converge com o diagnóstico de Viberti. Sobre a venda do TodoBr para o Google, já em 2005, ou seja, num momento posterior ao acima mencionado, Soares afirma:

“Esse era um caminho, digamos mais ou menos esperado em algum momento. Porque é aquele negócio. Você tem uma vendinha na vizinhança da sua casa, uma hora o Carrefour vai chegar lá e vai comprar. Porque você está tirando o mercado do Carrefour. Então a mesma coisa aconteceu com a gente. A gente tinha o mercado brasileiro de sistema de busca, porque o *TodoBR* em si ele nunca foi um sistema muito popular, mas entre 2000 e 2004, a gente fornecia tecnologia para o UOL, para o IG, para o IBEF, pra vários outros sistemas populares. Estão todo mundo estava usando esses sistemas, na realidade estavam usando tecnologia do *TodoBR*. E a empresa se mantinha inclusive a partir desses contratos. (...) Então a gente começou a incomodar o Google, que estava crescendo, querendo crescer nos mercados fora dos Estados Unidos, como na América Latina e na Ásia. E a gente tinha plano de expansão na América Latina inteira, isso era uma coisa que estava na pauta. A venda para o Google se deu porque no fundo, a proposta era muito boa e não tínhamos muito como sustentar, eu acho, aquela outra perspectiva. Como sustentar uma máquina de busca puramente brasileira sem investimento deles?” (SOARES, 2014).

Embora não seja o objetivo deste estudo refletir sobre as implicações do mercado de internet no Brasil, ou mesmo, os desafios apresentados para o desenvolvimento de empresas nacionais de ponta, não podemos deixar de apontar a relevância de tais questões. Ao tratar neste item um pouco da trajetória do Cadê estamos indicando alguns dos possíveis usos da coleção digital. Penasmo que a investigação das interfaces das ferramentas de modo comparativo a partir das fontes nascidas digitais, depositadas no Buscas.br, pode ser um caminho para compreender as singularidades da web 1.0 no Brasil e tentar compreender melhor como se deram os processos de adoção de novas tecnologias e as disputas econômicas a elas associadas.

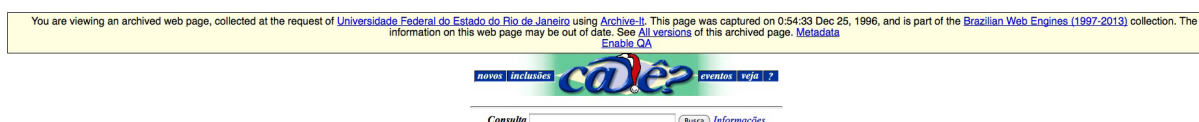
4.3 Em busca das buscas

“A internet (e tudo isso que estamos chamando aqui de ciberespaço) é como se fosse um universo paralelo – só que inteiramente eletrônico. Algumas pessoas perguntam: mas onde fica a internet? Ora, não fica. Não existe uma coisa física chamada internet. Você não pode vê-la, tocá-la ou ouvi-la.” Sérgio Charlab, *Você e a Internet no Brasil*, editora Objetiva, 1995, p.21. (CHARLAB,1995:21)

Ao propor o estudo das criptografias da memória da web brasileira nos propomos a fazer um contraponto com a imagem mitológica de um “ciberespaço” imaterial e, em algumas descrições, quase onírico. Esta pesquisa segue a trilha de vários pesquisadores de história das mídias (BURKE, 2008; GITELMAN, 2008) que propõem justamente a necessidade de se nomear os agentes sociais, os fatores culturais e as materialidades envolvidas nas diversas configurações tecnológicas.

Assim, ao selecionar e analisar a coleção foi possível problematizar alguns postulados, que pensados de maneira mais geral ao vasto campo, que inclui desde documentos digitalizados até a produção de memória em plataformas da web 2.0, pudemos perceber que tal fluidez não se adequa ao passado arquivado no Internet Archive. Algumas questões se apresentaram, como, por exemplo, o que é comumente chamado de incoerência temporal. Ou seja, quando a data do arquivo não está correspondendo ao conteúdo, ou a parte do conteúdo, já que a composição da página arquivada pode se dar de modos diferentes a cada atualização. A imagem abaixo fornece um exemplo, a mensagem escrita diz: You are viewing an archived page, collected at request of Universidade Federal do Estado do Rio de Janeiro using Archive-It. This page was captured on 0:54:33 Dec 25, 1996, and is part of the Brazilian Web Engines (1997-2013)”:

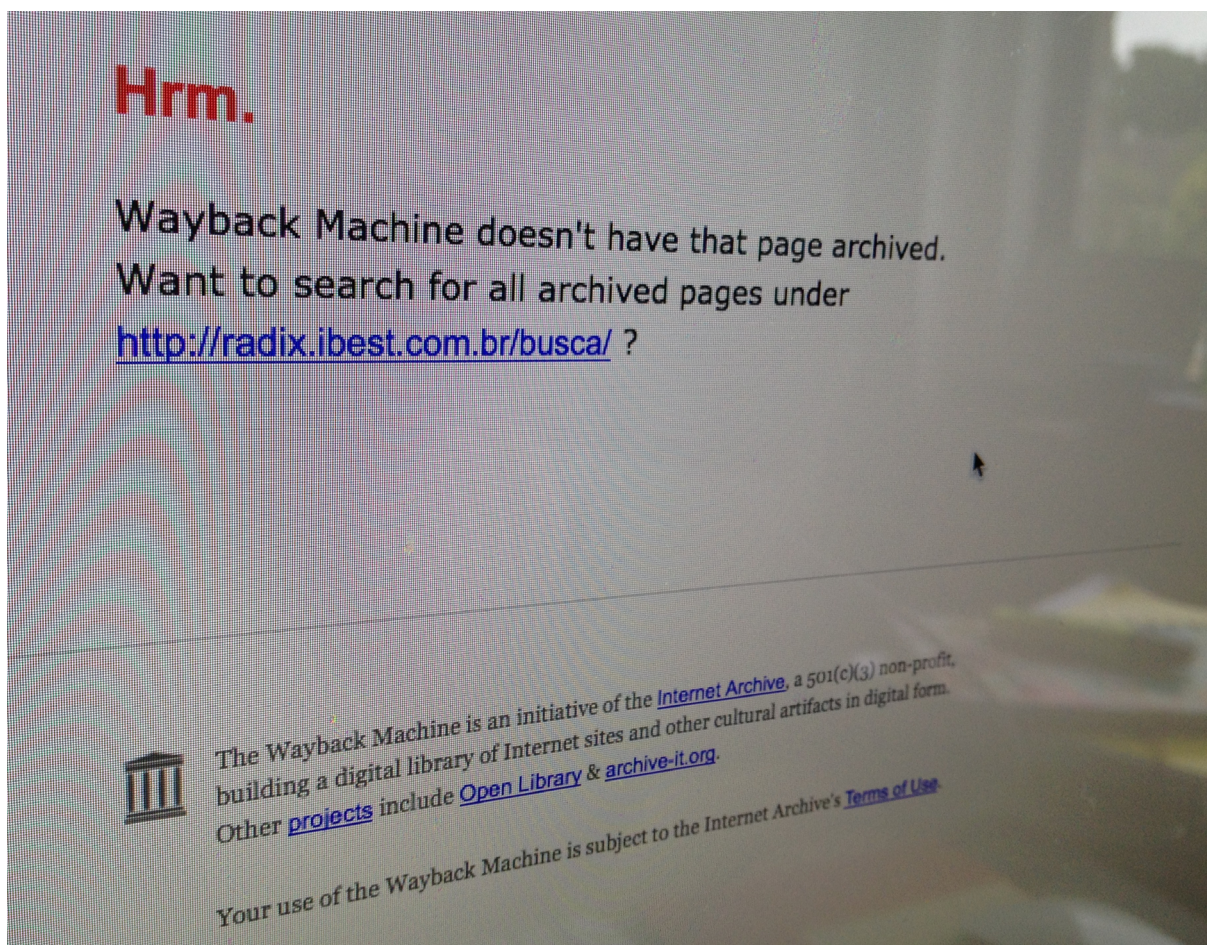
Imagem 41 - Incoerência temporal no registro do website *Cadê?* no *Internet Archive*



Por exemplo, uma das páginas incluídas na coleção, registrada na linha do tempo como tendo sido criada em 1997 foi capturada em dezembro de 1996. Este é apenas um aspecto onde há uma sobreposição de temporalidades em um documento que possui um selo temporal (time stamp) único. Esta é certamente uma questão a refletir, sobre as qualidades próprias das fontes primárias digitais.

Como mencionamos no capítulo anterior, Wolfgang Ernst, em *Archives in Transition* (2013) argumenta que há uma ruptura na estética do armazenamento, no que ele classifica de memória dinâmica. Os arquivos digitais representam um intervalo entre a memória e o presente, mas que têm trazido o presente cada vez mais para perto daquilo que é armazenando provocando um curto-circuito da memória (2013:99). No sentido de uma superposição das temporalidades, o termo memória dinâmica é de fato adequado para descrever, por exemplo, o fenômeno acima citado. Há de fato um movimento, uma performance na consulta ao arquivo, parte constitutiva da materialidade digital (Manovich, 2008). Porém, ao elaborarmos e consultarmos esta coleção percebemos que a dinâmica não diz respeito a uma interação simétrica entre usuários e plataforma/coleção. Na verdade as contingências técnicas muitas vezes nos levam a telas vazias, a longos tempos de espera que em nada se assemelham ao que comumente poderíamos associar a uma memória dinâmica. As telas abaixo mostram algumas falhas do sistema que nos ajudaram a pensar sobre a noção de criptografias da memória.

Imagem 42 - Falha no arquivamento do website *Radix*



Além de fazer a captura da tela via o próprio software, no meu caso o o utilitário Apple Grab, que permite a seleção de parte da tela, optei por fotografar a imagem do arquivo em consulta. O objetivo foi o de tentar revelar como o olho humano percebe a página e deixar a mostra os reflexos na tela. A textura revela também um pouco da poeira digital que é retirada nas imagens automáticas, tipo print screen. Outra forma de perceber aspectos criptográficos é focar no Glitch (ruídos ou falhas do sistema) apresentados nas sessões de consulta.

A importância do Glitch, como parte integrante do sistema, é justamente dar a ver a materialidade do registro digital. Neste momento gostaríamos de tão somente apontar para uma vereda a ser perseguida no futuro, qual seja, o gosto de glitch do arquivo digital⁸⁶. A arte do Glitch com seus manifestos e elaboração de aplicativos para criação de imagens com falhas é um fenômeno importante do início desta década do século XXI. A estética do Glitch⁸⁷, segundo a artista Rosa Menckman (2010) tem por objetivo estabelecer um contato criativo com as falhas dos sistemas, rompendo com uma suposta transparência dos protocolos e o padrão de invisibilidade dos padrões técnicos que informam as nossas práticas cotidianas. A proposta dos grupos que trabalham com esta estética é provocar o espectador; tensionar as mediações entre o olho e a tela, entre o comando e o resultado na inscrição.

Inspirada nesta proposta artística, gostaria de propor que a coleção ora apresentada, embora bastante incompleta em seu escopo documental sobre a web brasileira, possa ao menos contribuir no sentido de dar a ver as ruínas e provocar novas elaborações. Iniciamos nossa pesquisa analisando os próprios objetos digitais, seguindo a perspectiva de Richard Rogers (2008) de estudar os objetos nascidos digitais a partir de suas características nativas. Esta escolha nos fez mergulhar na materialidade dos registros digitais, ou seja, nas páginas arquivadas da web. Investigar a memória social através de um experimento prático nos possibilitou repensar alguns postulados referentes à memória digital. Há um discurso hegemônico que confunde a característica de transcodificação própria aos objetos digitais (por meio da linguagem numérica) a uma suposta flexibilidade da memória digital que, muitas vezes inclusive, acaba sendo atribuída aos repositórios digitais⁸⁸. Tal abordagem pode levar a assertivas por demais generalizantes e, em minha perspectiva, falsas, sobre a natureza dos arquivos digitais. Este é o caso da seguinte afirmação: “Digital archives are unstable, plastic, living entities, as stories and rituals were in oral cultures.” (BROWER and MULDER, 2003:5).

86 O gosto aqui remete, novamente, ao trabalho pioneiro de Arlette Farge (1989)

87 Disponível em: <http://rosa-menckman.blogspot.nl/> (acesso em 10/05/2014)

88 Um exemplo deste discurso está na seguinte passagem: “A memória digital está mais próxima à memória da oralidade que à escrita, uma vez que a transmissão de conhecimentos entre os indivíduos de um grupo ocorre de maneira semelhante: em tempo real. (MONTEIRO, CARELLI: 2007:18)

Na verdade, embora seja preciso admitir a instabilidade dos arquivos digitais, não identificamos no processo de elaboração desta coleção digital nada que possa indicar uma semelhança importante com os modos de produção e recriação da cultura oral. As especificidades técnicas demonstraram uma fixidez nos registros e as lacunas assemelham-se, em grande medida, àquelas encontradas em arquivos físicos. Ao mesmo tempo, alguns aspectos técnicos, sobretudo no que se refere a uma memória documental, podem significar uma restrição do acesso aos documentos. É evidente que os arquivos tradicionais também são espaços restritos e que exigem conhecimentos específicos como a escrita, mas as novas tecnologias passam a demandar algumas outras habilidades para acessar os dados. Muitas vezes é preciso ter, por exemplo, uma mínima habilidade no uso de alguns formatos digitais e uma compreensão do que eles representam.

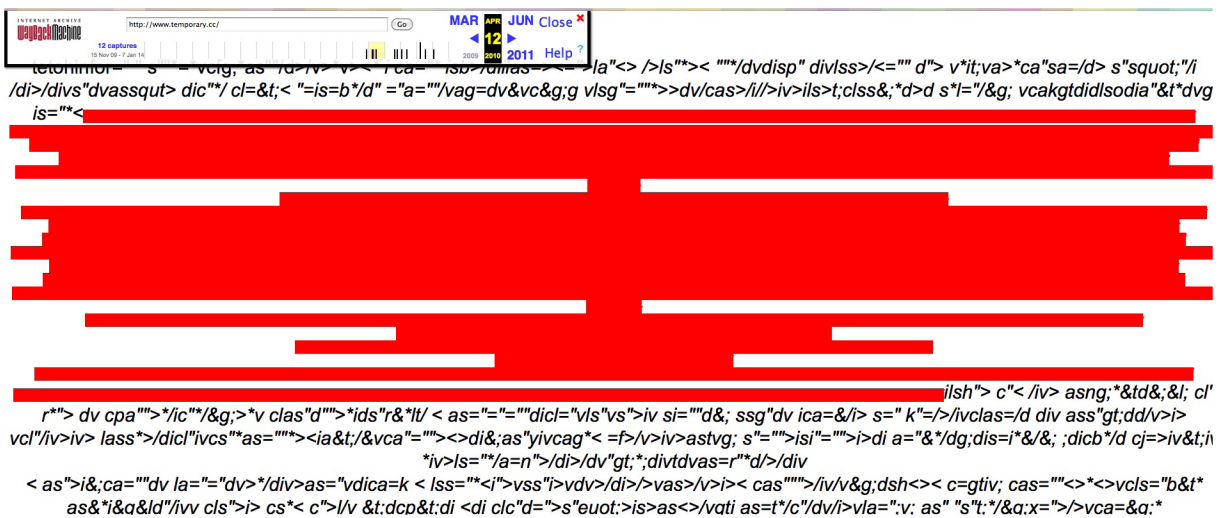
Por tudo isso, não podemos concordar com afirmações genéricas de que a memória digital é mais acessível ou similar à memória oral e que aconteceria ainda em tempo real (MONTEIRO, CARELLI, 2007:7). Os postulados desta natureza ignoram os processos concretos de circulação dos registros digitais. Tal hipótese está associada a uma perspectiva que toma as tecnologias em si como elementos de mudanças sociais. A realização de uma investigação prática, nos fornece evidências para afirmar que o caráter dinâmico da página arquivada não significa uma fluidez no que se refere a construção ou acesso aos fragmentos de uma memória documental nascida digital. Se há algo específico e que merece ser investigado, nos parece ser uma estética Glitch que perpassa a coleção. Neste sentido, as criptografias da memória ali contidas, podem engendrar novos glitches, ou pelo menos, essa é a aposta. Concluo este capítulo contrapondo um trecho do manifesto Glitch (Menckman:2010) com uma imagem glitch da própria coleção Buscas.Br.

Imagem 43 - Trecho do manifesto Glitch, de Menckman (2010)

the glitch is a momentary experience of an interruption that shifts an object away from its ordinary form and discourse. For a moment I am shocked, lost and in awe, asking myself what this other utterance is, how was it created. Is it perhaps ...a glitch? But once I named it, the momentum -the glitch- is no more...

ut somewhere within the destructed ruins of meaning exists; a triumphal sensation that there is something more than just devastation. The negative feelings make place for an intimate, personal experience of a machine (or program), a system showing its inner workings and flaws. As a celebration rather than a particular perfection a glitch can reveal a new opportunity, a spark of creativity that indicates that something new is about to be created.

Imagem 44 - Imagem glitch da coleção *Buscas.Br*



Fonte: Internet Archive

CONSIDERAÇÕES FINAIS

A investigação das criptografias da memória nos conduziu a um ponto de vista que diverge das perspectivas que associam a tecnologia digital e as memórias produzidas neste suporte a noções como as de acessibilidade e democratização da informação. A coleção Buscas.br pode, certamente, ser considerada um passo no sentido de uma apropriação criativa das ruínas digitais. No entanto, os percalços encontrados e os limites do próprio experimento nos levam a perceber uma série de entraves, culturais e tecnológicos, relativos a preservação e ao acesso à memória digital.

No contexto brasileiro vimos que os registros nascidos digitais não têm sido alvo de uma política ativa de preservação. Como exposto na discussão do Forum da Cultura Digital(2010), no que se refere à elaboração de políticas públicas que incluam a preservação da memória documental nascida digital, ainda esbarramos em conceituações calcadas no patrimônio físico, daí a ênfase, pensamos, na digitalização de acervos. A criação da Rede Brasileira de Serviços de Preservação Digital, em 2013, indica algumas mudanças neste cenário embora os principais objetivos desta iniciativa estejam relacionados a preservação do conhecimento acadêmico.

Considero a experimentação prática envolvendo concepção, curadoria e realização da coleção Buscas.Br como o principal resultado deste trabalho de pesquisa. O processo de construir a coleção nos deu a possibilidade de problematizar nossas reflexões teóricas e aprofundá-las em outras direções. Como os registros digitais poderiam se tornar memória? A relevância da investigação teórica aqui proposta, sobretudo na enunciação da noção de criptografias da memória, está justamente no diálogo proposto com os processos de elaboração da memória cultural produzida e acessada em suporte digital. Foi a partir de um primeiro mergulho em um conjunto de referências sobre as materialidades dos códigos e dígitos, seguindo a trilha da História Cultural e dos Estudos de Software, que pudemos criar a possibilidade investigativa que tornaria esta pesquisa de doutorado um experimento.

A coleção Buscas.br, embora seja um ponto minúsculo no universo documental da web brasileira, pretende-se mais ampla no sentido de representar a incorporação de objetos que podem ser considerados um certo lixo cultural. As páginas arquivadas, ou seja, as ruínas em WARC, não deixam de propor um olhar justamente para o lado mundano, comercial e pouco glorioso da web. Não se trata de um acervo já institucionalizado ou considerado relevante por algum grupo cultural. São fragmentos dos primeiros anos da web no Brasil, suas formas de classificação e maneiras de acessar. A coleção talvez aproxime-se mais de um repositório do esquecimento que de um arquivo de memórias. Pensando nos termos propostos por Aleida Assman (2011), destacamos o caráter de arquivo da coleção que, embora pública e acessível, encontra-se na reserva técnica da

web. O canon talvez se apresente nos primeiros resultados de busca do Google, cujas criptografias não podem ser livremente acessadas.

Gostaria de indicar nesta última parte futuros desdobramentos que não seriam possíveis desenvolver no âmbito deste projeto, mas que podem vir a se concretizar a partir da coleção Buscas.Br. Uma aspiração é tornar a coleção mais dinâmica, um pouco mais próxima daquilo que Huyssen (1986) e Ernest (1999) descreveram ao conceber os bancos de memória computadorizada. Neste sentido, pensei em dois desdobramentos. O primeiro seria a possibilidade de inclusão da documentação oral e escrita, recolhida ao longo da pesquisa, porém não plenamente incorporada neste estudo. Embora o estudo tenha partido da necessidade de perceber as materialidades digitais, a realização da coleção apontou para as lacunas deste tipo de documentação. Assim, é factível pensar numa plataforma que possibilite a inclusão de diversos tipos de mídia e favoreça uma leitura sincrônica da documentação sobre o passado da web no Brasil.

Um outro desdobramento interessante seria abrir a coleção para que o público também pudesse contribuir na inclusão de metadados. Neste caso, isto envolveria a migração para outra plataforma, que seja mais amigável. Embora possa parecer um desdobramento um tanto arriscado, é importante salientar que já existem experiências do gênero voltadas para a inclusão de categorias em imagens (crowdsourcing).

Um aspecto a ser salientado é que a arquitetura da coleção insere-se no campo de convergência de perspectivas de preservação e divulgação do legado cultural. A construção de uma coleção em movimento contínuo poderia aproximar esse experimento da tese da ideia de convergência entre as funções do arquivo, da biblioteca e do museu (ROBINSON,2010). Aproximar apenas em alguma medida, já que entendemos que as fronteiras, embora possam se tornar tênues, ainda existem e as especificidades podem por vezes ser benéficas. A construção de uma coleção em movimento contínuo poderia aproximar esse experimento da tese da idéia de convergência entre as funções do arquivo, da biblioteca e do museu. Aproximar apenas em alguma medida, já que entendemos que as fronteiras, embora possam se tornar tênues, ainda existem e as especificidades podem por vezes ser benéficas.

A inscrição da coleção Buscas.Br num circuito mais amplo, inclusive considerando a possibilidade de sua utilização pedagógica, pode ser um modo de contribuir para aquilo que Beatriz Sarlo (2008) denominou de construção de uma guinada crítica da memória. Promover o acesso aos vestígios nascidos digitais pode ser uma forma de tornar visível a historicidade das tecnologias da informação, mais especificamente da web. No caso do passado da web, seria importante agregar a possibilidade de interrogar uma memória documental, sem seguir as narrativas utópicas ou alarmistas. Dar a ver os ruídos, o *glitch* na tela, é sair do fluxo corrente de navegação contínua; é um gesto no sentido de promover um olhar crítico sobre as variadas criptografias da memória digital.

ANEXOS I

*As entrevistas a seguir foram editadas no sentido de manter clareza,
retirando ruídos da comunicação oral.*

Entrevista com Altigram Soares, via Skype, em 30 de junho de 2014.

C: Bom dia, *Altigram*. Obrigada mais uma vez por participar nesta pesquisa de doutorado. Hoje é dia 30 de Junho de 2014, e vou entrevistar *Altigram Soares*, cientista da computação e professor da *Universidade Federal do Amazonas*. Minha primeira pergunta é um pedido de apresentação, fale sobre a sua trajetória profissional e como foi sua experiência inicial no *TodoBR*.

A: Bom, eu sou professor aqui na UFAM há bastante tempo, vinte e poucos anos. E fiz meu mestrado e doutorado na *Universidade Federal de Minas Gerais*. O *TodoBR* começou na época em que eu estava fazendo o doutorado na UFMG. Os meus coordenadores, professor *Alberto Lãnder* e professor *Berthier Ribeiro Neto*, na época, eles iniciaram esse projeto junto com o professor *Nívio Ziviani*, e eu recebi o convite para participar por conta das atividades que eu estava fazendo no meu doutorado. E assim foi como eu me envolvi no *TodoBR*.

C: Você entrou no *TodoBR* nesse início, em 1999?

A: Foi. Na verdade, eu fui um dos primeiros a entrar no projeto. Justamente em 1999, a convite do *Berthier*, eu comecei a desenvolver os códigos do *TodoBR*.

C: Qual era a sua área específica de trabalho nessa equipe? Poderia nos falar um pouco sobre isso?

A: Posso. Eu, na época, precisava para o meu doutorado de coletar dados da *web* para fazer experimentos, para isso, eu escrevi um *coletor*, um *crawler*. E aí, meu trabalho nessa época, bem no começo era no *crawler*. Eu acho que escrevi um dos primeiros *crawlers* da *web brasileira* na época. O que eu fiz foi expandir esse projetinho pequeno que eu tinha, que era algo do doutorado, e acabei escrevendo o primeiro *crawler* do *TodoBR*. O *crawler* é um componente da máquina de busca, que coleta as páginas para serem armazenadas e depois indexadas. Então, na época trabalhavam comigo dois alunos de graduação, e juntos escrevemos o primeiro *crawler* do *TodoBR*.

C: Esse *crawler* especificamente só indexava o domínio “.br”?

A: Sim. Porque nosso objetivo era indexar páginas da *web brasileira*. Por isso, o nome da maquininha era *TodoBR*, justamente para cobrir bem o Brasil. A gente sabia já naquela época, que Google tinha entrado no ar um pouquinho antes, no ano de 1998 e na época eu estava escrevendo o *crawler*. E o *Google* tinha a melhor cobertura entre as máquinas de busca do mundo. Mas a gente sabia, e eu cheguei a fazer experimentos, que a cobertura dele para a *web brasileira* era muito pequena. E foi justamente aí que a gente resolveu atacar. Então nessa época, nós éramos melhores do que o Google na cobertura da *web brasileira*, pois cobríamos muito bem a *web brasileira*. A gente estava no backbone da RNP. Naquele tempo, o backbone da RNP era

o principal backbone que tinha e cobria basicamente todo mundo. Tinha também uma troca de tráfego boa com a EMBRATEL. E com isso, nós podíamos cobrir bastante a *web brasileira*. Nossa cobertura era bem melhor.

C: Você lembra que nessa época, creio que até 1999, a RNP ainda tinha um buscador chamado *Yahi?*, que era semelhante ao modelo dos diretórios do Yahoo?

A: Não, realmente não lembro não. O que eu lembro era do Cadê.

C: Sim, o Cadê? Também tinha um modelo diferente de diretórios, mas em termos de utilização pelos usuários na época, de 1998 a 1999, ele era o principal portal de busca.

A: Sim

C: Vocês tinham contatos com o Cadê ou o Radix, em Pernambuco, que também estavam começando?

A: Então. O Cadê a gente sabia basicamente por causa da imprensa. O que você tem que entender, é que o *TodoBR*, inicialmente, ele era um experimento puramente acadêmico. A ideia era colocar vários trabalhos de pesquisa que estavam sendo feitos naquela época lá do DCC da UFMG, em prática. O *TodoBR* era um laboratório, ele no início não tinha pretensão de ser um produto. O Radix a gente sabia mais por causa do envolvimento com a comunidade acadêmica. A gente conhecia o pessoal lá da UFPE, algumas pessoas como o *Sílvio*, a *Carol*, mas não chegamos a trocar idéias sobre o que fazer. E o pessoal do Cadê, eles muito cedo conseguiram investimento, então tinha-se muita exposição na mídia e muitos usuários. Os nossos usuários apareceram bem mais tarde, quando a gente resolveu colocar a máquina de busca no ar. Mas por outro lado, nós tínhamos um controle muito grande do processo inteiro, porque nós desenvolvemos todo o software do zero.

C: Esse *crawler* que você desenvolveu, pioneiro no Brasil, além de ser programado para o domínio “.br”, tinha filtros. Quais eram os principais filtros de retorno das consultas dos usuários?

A: Deixa eu ver se eu te explico. O *TodoBR*, não é que ele fosse específico para o o domínio “.br”, mas ele foi preparado para coletar países. E depois a gente levou o *TodoBR* para o Chile e Espanha. E a gente não mudou absolutamente nada no código. A gente levou pra lá e rodou. É basicamente isso. E ele usava para isso a fronteira, vamos dizer assim, geofísica dos IP's. A gente conhecia o mapa dos IP's, e usava isso para não deixar o *crawler* vazar para fora do país.



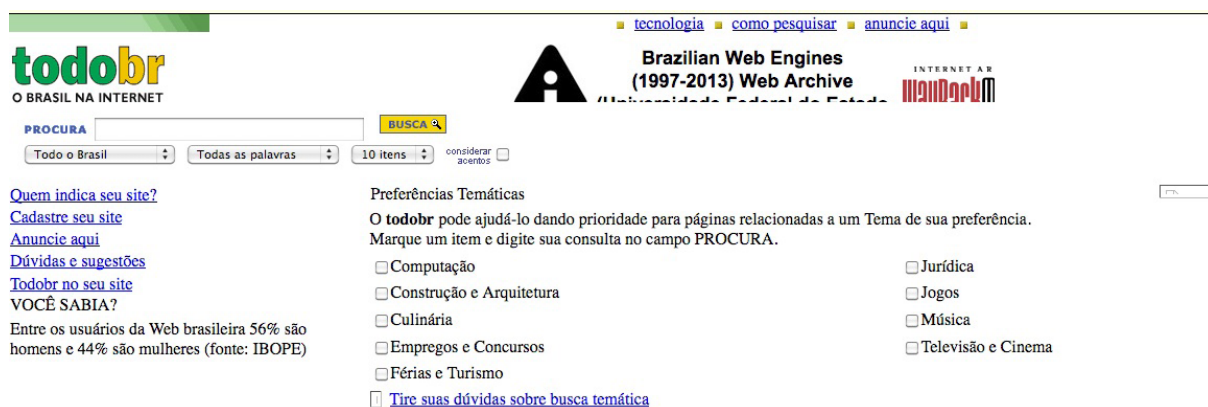
Página arquivada pelo Internet Archive, 28/11/1999.

C: Então vocês não programavam para ter o domínio “.br”? Poderia ser um site alojado no Brasil “.com” ou “.gov”, sem o domínio “.br”? Por que na verdade o “.com” ele era alojado em servidores fora?

A: Exato.

C: Mas como, por exemplo, uma página em português do Brasil que tivesse o domínio “.com” era encontrada pelo *crawler*?

A: Ele pegava, porque a gente usava na realidade o mapa de IP’s, que o link “.br” deixava isso disponível. Então a gente mapeava através do mapa de IP’s. A gente usava alguns casinhos “.br”, mas o grosso era através do mapa de IP’s. Então através de uma página de hospedagem do Brasil, com seguidores de dentro do Brasil, a gente ia lá e pegava. Inclusive, a gente sabia a localização física das páginas. O *TodoBR* chegou a ter bastante tempo que você poderia fazer, por exemplo, páginas que estão em Pernambuco, ou que estão em Minas Gerais, ou no Rio, a gente conseguia fazer isso.



Página arquivada no Internet Archive, 5/01/2001.

C.: Eu pesquisei no *internet archive* e vi que a caixa de busca de vocês tinha essa possibilidade. Um dos objetivos da minha pesquisa é recuperar essas páginas como eram e disponibilizá-las para os usuários quando meu projeto ficar pronto, por isso estou tentando entender essas características.

A: Hoje não funcionaria. Porque hoje você tem uma falta muito grande de soluções para trocar tráfegos e naquela época as coisas eram bem mais, digamos, controláveis. Hoje em dia, talvez não funcionasse, mas na época funcionou muito bem.

C: Nesses estados nos quais vocês trabalhavam, as pessoas começaram a mandar e-mail para se inscreverem ou a maior parte da indexação foi feita via o *crawler*?

A: Não, a indexação era toda feita via o *crawler*.

C: Pergunto isso porque nos sites tem a seguinte mensagem: se você quer inscrever sua página, mande sua página.

A: É, isso foi muito depois. Mais aí a ideia já era outra. A ideia era a seguinte. A *web*, até hoje é assim. Você tem regiões que são encobertas. Então pra chegar lá, o *crawler* ou tem que andar muito, ele então não vai chegar a lugar algum, porque você não tem apontador pra lá. Naquela época, ainda era menos denso. Então, se você, por exemplo, me desse o link da sua página, eu não só colocava a sua página, como eu a partir dela iria encontrar uma região da *web* que não estava coberta. Teve uma época que nós coletávamos o *bookmarklet* das pessoas, e as pessoas também mandavam os *bookmarklets*. E a partir dos *bookmarklets* a gente conseguia aumentar a cobertura do *crawler*. O *crawler* ele trabalha com semente. Então na realidade essas 10:09 [ruído - palavra não identificada] serviam como semente. Era mais do que simplesmente se você dissesse assim: me dê sua página. Mas eu seguramente posso te falar que 95% de tudo que a gente tinha na base, o *crawler* descobria.

C: Você está falando que o *crawler* tinha justamente o objetivo de mapear essa *deep web*, que ainda não estava mapeada, é isso?

A: Isso.

C.: Vocês pegaram esse mapa de IP e fizeram um mapeamento inicial desses primeiros *seeds*, as primeiras sementes dos endereços eletrônicos URL, ou fizeram de outra forma?

A: A gente, nesta época, começou a sair as primeiras páginas do ODP, *Open Directory Project*. Então a gente começou a filtrar as URL's do “.br” que vinham e estavam nesse diretório. E ai deu um conjunto muito bom, com uma cobertura muito boa. A gente na época tinha uma regra de ouro, que a *web brasileira* era 1% da web mundial. Agora já mudou, mas se valeu por vários anos. A gente tentava estimar essa cobertura por base justamente dos dados que vinham do Google, aquele famoso artigo da *net*. Então a gente sabia que a cobertura estava boa

por causa disso. Então a ideia era aumentar essa cobertura e aumentar também a densidade de links, que eu sempre achei muito baixa. Então a gente tendo mais URL's a gente conseguia medir a densidade de links e conseguia saber se o *crawler* estava indo bem ou mal.

C: Nessa época, com a base de dados e os meios que dispunham, vocês chegaram a fazer visualizações da *web brasileira*?

A: A gente fez tem um artigo, acho que foi publicado na WWW. Eu vou demorar a achar agora. O pessoal fez uma visualização. Na realidade, o pessoal do Chile, nossos parceiros no Chile, o *Ricardo Baeza-Yates* fez primeiro que a gente. E ai depois eu vi o trabalho dele e pedi o software, e ai a gente fez da *web brasileira*. Mas eu sinceremante não sei aonde que isso foi parar.

C: Como essa pesquisa é sobre história e memória, eu vou fazer perguntas sobre aquela época. O *TodoBR* então nasceu na UFMG e depois em 2002 ele passou a ser uma empresa, o *TodoBR SA, correto?*

A: Sim.

C: E nesse momento você foi para empresa ou ficou como pesquisador da Universidade? Essa pergunta é para que possamos conhecer melhor a sua trajetória.

A: Eu era doutorando e eu sempre trabalhei aqui na UFAM. Quando eu fui fazer o doutorado, eu fui vinculado a UFAM. Eu não podia ir para a empresa. O que eu fiquei fazendo foi basicamente trabalhando com os meninos lá da *Akwan*. Eu orientava eles a fazerem o *crawler* de lá. Eu não tinha um envolvimento muito direto com a *Akwan*. Isso era um negócio mais tênuo. Eu interagia com o pessoal, indo lá assim, mostrando como que tinha que ser feito, dando ideias, a gente discutia e debatia muito as coisas, mas não tinha um envolvimento muito direto com a empresa. Justamente porque eu era professor aqui da Universidade, então não tinha muito como eu me vincular.

C: Você ainda tem a base de dados desse primeiro *crawler* que você fez e foi sua pesquisa de doutorado?

A: Não, de jeito nenhum.

C: Você conhece algum projeto no Brasil dessa área na Ciência da Computação que tem o objetivo de preservar, sejam as páginas ou os códigos?

A: Eu realmente não sei, porque talvez a gente na época não tenha tido muita noção da importância histórica pro país. E não tínhamos mesmo essa noção. A gente na época, era meio amador. Talvez hoje eu até tivesse essa preocupação, mas na época não.

C: O objetivo inicial do meu trabalho era verificar como seria pesquisar a história da *web brasileira*; como se poderia para o historiador estudar documentos eletrônicos. Comecei pes-

quisando no *internet archive* e entre uma coisa e outra resolvi pesquisar especificamente sobre as ferramentas de buscas ou engenhos de busca, como eram chamados na época. Foi então que constatei que só existe uma parte disponível no *internet archive*. Portugal já tem um web archive bem interessante, ligado ao grupo da Ciência da Informação, mas no Brasil desconheço. Entrevistei algumas pessoas, pois poderia ter alguma iniciativa dentro das Universidades que não foram amplamente divulgadas.

A: Eu realmente não sei.

C: Bom saber, essa resposta é importante para mim.

A: Agora, eu estou pensando aqui o seguinte. Eu acho, para te falar a verdade, tem uma base que deve ser de 2003 ou 2004, que o pessoal usa até hoje para fazer experimento. Essa base, eu acho que foi logo antes da ida pro Google, porque depois que fomos pro Google a gente não pode mais usar a base. O *Edleno Moura*, não sei se você passou por ele aí quando você estava estudando. Mas o *Edleno*, hoje é meu colega aqui na UFAM, mas ele foi o CEO da *Akwan*.

C: Já vi o nome dele nas histórias mas ainda não o entrevistei.

A: O processador de consultas do *TodoBR* foi ele que escreveu, ele com outras pessoas. Mas a iniciativa partiu dele. Ele teve um maior envolvimento com a *Akwan*, porque ele ficou como o CTO até um pouquinho antes da venda pro Google. Ele saiu um pouco antes, acho que na verdade foi em 2003. Mas eu acho que o *Edleno* tem hoje essas bases que são usadas para fazer experimentos. Ele usa até hoje na pesquisa dele. O *Edleno* é um dos principais estudantes da área de indexação do mundo. Eu acho que ele usa essa base até hoje.

C: Então nessa época você continuou pesquisando. A *Akwan* continuou em Minas e em 2002 foi vendida pro Google?

A: Foi em 2005.

C: Eu gostaria de saber do que você lembra daquele momento, do mercado naquele período e mesmo antes. Você se recorda de outras iniciativas que não deslancharam? Pergunto porque o Radix acabou sendo vendido, assim como o Cadê foi para o Yahoo e o *TodoBR* foi para o Google.

A: Esse era um caminho, digamos mais ou menos esperado em algum momento. Porque é aquele negócio. Você tem uma vendinha na vizinhança da sua casa, uma hora o Carrefour vai chegar lá e vai comprar. Porque você está tirando o mercado do Carrefour. Então a mesma coisa aconteceu com a gente. A gente tinha o mercado brasileiro de sistema de busca, porque o *TodoBR* em si ele nunca foi um sistema muito popular, mas entre 2000 e 2004, a gente fornecia tecnologia para o UOL, para o IG, para o IBEF, pra vários outros sistemas populares. Estão todo mundo estava usando esses sistemas, na realidade estavam usando tecnologia do *TodoBR*. E a

empresa se mantinha inclusive a partir desses contratos. A gente também forneceu no Chile, que lá tinha o *TodoCL*, que era a mesma base e sistema do *TodoBR*. Lá na Espanha, a gente tinha o buscador do próprio El país, que também usava a nossa tecnologia licenciada para eles, que era um dos maiores clientes na época. Então a gente começou a incomodar o Google que estava crescendo, querendo crescer nos mercados fora dos Estados Unidos, como na América Latina e na Ásia. E a gente tinha plano de expansão na América Latina inteira, isso era uma coisa que estava na pauta. E aí, a venda pro Google se deu porque no fundo, a proposta era muito boa e não tínhamos muito como sustentar, eu acho, aquela outra perspectiva. Como sustentar uma máquina de busca puramente brasileira sem investimento deles.

C: Será?!

A: A venda foi uma coisa boa, eu acho.

C: Hoje em dia estamos falando e já sabendo o que o Google é; o Google no Brasil tem hoje mais de 90% de alcance das buscas, mas mesmo naquela época inicial já havia entre os profissionais da área essa visão de que o Google já era esse, digamos, ‘super poder’?

A: Isso iria acontecer fatalmente. Bom, sim. Porque tinha-se uma boa tecnologia, as pessoas tinham um preparo muito bom, um pessoal que estava a frente do Google, mas os investimentos que eles estavam tendo e a atenção que eles tinham do mercado era impressionante, já naquela época. Então isso iria acontecer de um jeito ou de outro.

C: Essa é sua visão? Na época não havia espaço, por exemplo, para um empresa brasileira nessa área?

A: É. Essa é minha visão, embora eu não tenha tido uma participação assim, mais efetiva da decisão.

C: O importante para mim é saber as visões da época. Porque uma coisa é hoje, que nós falamos e já sabemos o que é o Google, mas eu gostaria de saber como era naquele momento.

A: É um pouco como se você quisesse ter uma montadora de carros brasileira. Quero dizer, é muito difícil, embora tenha mercado, tecnologia e gente capacitada, não se achava. Não imagino isso com um fabricante de computadores, por exemplo, com tecnologia própria. Era muito difícil. Isso aconteceu no mercado de aeronaves, tendo um investimento pesado do próprio governo.

C: Naquela época o governo não tinha interesse em fazer grandes investimentos, através da RNP, por exemplo? Houve investimento governamental enquanto sendo um projeto da Universidade e das pesquisas, mas estou questionando sobre manter o investimento e aprofundá-lo.

A: Sinceramente, eu não acho não, *Camila*. Eu acho que historicamente o investimento que o Governo direcionou para ciência e tecnologia ele é muito voltado para o momento. Naque-

la época era um projeto visionário. Então, imagina que na época alguém poderia parar e dizer: Ah, pô mais, será que daqui a 20 anos a gente vai ter uma tecnologia forte nessa área? Eu acho que não. Na época se investia sempre em condutor, para você ter uma idéia.

C: Mas por um lado é estranho porque você desenvolveu um tipo de *crawler* que era tão bom, na época, quanto os resultados do Google!

A: É, pois é! Assim, veja só. De novo, o *TodoBR* foi desenvolvido para ser um laboratório. Então na época já achavam que tecnologia que tínhamos, não só o *crawler*, mas o processador de consultas e várias outras coisas que a gente tinha. Por exemplo, o processador de consultas nosso, ele rodava em um PC, para você ter uma ideia. Depois que a gente colocou um outro, mas era de redundância, porque o papel do *Edleno* na época era de compressão de dados, era a melhor técnica. Foi durante muitos anos a melhor técnica de compressão de dados que existia na academia. E a gente usou ela pesadamente no *TodoBR*. Então, não só o *crawler*, mas em várias outras frentes, tecnologicamente o *TodoBR* era muito avançado. A gente fazia na época, uma coisa que eu faço até hoje, que é pegar resultado direto das teses e aplicar pra colocar as pessoas para usarem. A gente tem outra *startup* que faz exatamente isso.

C: Mas é isso que estou falando. Você falou: de fato, nasceu na Universidade, mas, por outro lado, o Google também.

A: É, mas.

C: Eu sei que tem outra proporção, mas também foi um projeto de doutorado do *Larry Page* e do *Sergey Brin*.

A: É, mas é um outro contexto, completamente diferente.

C: Claro. Lá a tecnologia já estava mais desenvolvida, o acesso à tecnologia nas Universidades era muito maior.

A: E veja só. A gente está falando aqui, 15 anos depois, de uma *startup*, uma das primeiras *startups* que apareceram no país e só a dois anos atrás que a UFMG, por exemplo, regulamentou o processo de criação de *startups*. Há dois anos atrás não tinha nada. Mas todo dia você tem aluno e professores recebendo investimento para fazer desenvolvimento e aqui a gente fica 27:21 [ruído] porque a gente fundou uma *startup*, que eu não posso passar na porta, senão estou quebrando meu DR.

C: Sério?

A: É.

C: Não creio! A Universidade que funda e ela não pode estabelecer uma parceria?

A: É super difícil.

C: Até mesmo para manter, utilizar base de dados para experimentos...

A: É muito difícil.

C: Nossa!

A: Muito difícil. A selva da legislação e dos advogados é um negócio que dá desânimo.

C: Eu tenho curiosidade para entender como funciona esses meandros entre tecnologia, universidade e sociedade, até para que isso fique um pouco mais visível para outras pessoas, para os debates em pautas.

A: A nossa *startup*, eu posso até te passar o link dela aqui.

C: Certo, eu quero.

A: Só para te dar uma perspectiva. Vou passar aqui pelo skype.

C: Ótimo.

A: A nossa *startup* é essa daqui. Não sei se você já comprou pela Americana, ou pelo Submarino. Já comprou nesses sites?

C: Já comprei há muito tempo, agora estou morando na Holanda, mas já comprei.

A: Então, a nossa *startup* é que fornece o sistema de busca e recomendação de produto para essas duas empresas e para muitas outras.

C: Nossa. Muito bom.

A: E a gente já tem hoje 25% do mercado brasileiro de e-commerce.

C: Isso é muito bom para uma...

A: Bom, a gente aprendeu isso lá na época do *TodoBR- Akwan*. Foi lá que a gente, eu e *Eduino*, entendeu que era possível desenvolver tecnologia de ponta pro mercado. Então, a gente continua fazendo isso até hoje. A gente aprendeu lá e essa é a terceira empresa que a gente fundou, depois da *Akwan*.

C: Nossa.

A: Só que é sempre muito difícil trabalhando em Universidade Federal. Pessoal da 29:34 [ruído - sigla não identificada] deve passar a mesma coisa. Unirio deve ter problema também.

C: Na Unirio nós já entramos em contato. Tem um contato com o qual queremos fazer um projeto com o pessoal da Informática, mas como meu doutorado está sendo feito em parceria com essa Universidade aqui na Holanda, a minha pesquisa é mais voltada para a área de Humanas. Adoraria ter mais colaboração com pessoas da informática, que pudessem fazer determinados experimentos com esses dados, já que nós não tem esse *know-how*.

C: Agora, uma pergunta sobre o *TodoBR*: eu cheguei a te perguntar sobre os filtros e gostaria de saber, caso você recorde, se tinha algum filtro, por exemplo, para a questão do retorno de busca, seja de conteúdo político, seja de conteúdo sexual; e uma segunda pergunta: vocês tinham alguma política de privacidade com os dados das pessoas que faziam as buscas?

A: Não, a gente não tinha filtros para nada. Nunca tivemos. E na época a gente não armazenava dados de usuários. Não tínhamos espaço para isso, enfim. Não deu tempo de desenvolver as coisas que precisavam para isso.

C: Entendi. E sobre o ranking? Quando vocês retornavam a busca, o algoritmo que vocês desenvolveram era baseado no mesmo princípio do Google, da questão dos in-links, da citação dos próprios endereços, ou tinha algum outro critério?

A: A gente usava a presença das keywords, dos in-links, dos gráficos de links. A gente usava um algoritmo diferente do Google na realidade, que era o HITS. E a gente também fazia, uma certa época, se o cara escolhesse na região geográfica, a gente filtrava, não filtrava, a gente ‘re-rankiava’ pela posição geográfica.

C: Esse algoritmo HITS era o mesmo utilizado pelo Radix. Ele era bastante difundido na época, ele era o principal? Para eu entender a opção por ele.

A: O HITS é mais antigo que o algoritmo do Google, que foi publicado tempos depois só. E o HITS tinha uma vantagem, era muito mais barato de processar porque ele usa evidências locais, o Google usava evidências globais. Aliás isso foi uma das coisas boas que o Google trouxe, mostrar que isso funcionava, porque as pessoas não acreditavam muito nisso. Então não se achava que precisaria de usar evidências globais para o algoritmo. Enfim, o Google mostrou que essa era uma opção boa também. Se bem que muita gente, depois, contestou esse resultado. Depois que ele se tornou popular, essa discussão deixou de fazer sentido.

C: Entendi. Vocês faziam testes na época, por exemplo, comparando o *TodoBR*, o Google e o Altavista?

A: Fazíamos. A gente, se eu não me engano, eu não sei se ainda tenho os relatórios que a gente fez naquela época. O *Edleno* talvez tenha. O nosso ranking era tão bom quanto para páginas do Brasil. E basicamente a razão era porque nossa cobertura era maior, então nosso resultado era sempre bem melhor.

C: Muito interessante, mas essa história é pouco conhecida.

A: É, eu acho que ninguém nunca contou essa história direito não.

C: Eu encontrei algumas reportagens; algumas no jornal da época têm muitos trabalhos científicos, mas na linguagem da pesquisa, com análises e experimentos. Tem os seus artigos, tem o artigo da *família Miner* e várias dissertações e trabalhos acadêmicos; mas não tem esta história divulgada para um público maior, como você está ajudando a construir agora, se tudo der certo.

C: Naquele momento, que tinha a liderança interna do *Nívio* e das pessoas que você já mencionou, vocês tinham contatos internacionais também? Como era essa relação um tanto quanto acadêmica?

A: Tinha bastante. O *Berthier*, ele veio da UCLA, o doutorado dele foi em 95. E ele veio com muitos contatos na área de *information retrieval*, e o *Nívio* tinha também bastante contatos. O principal contato era o *Ricardo Baeza-Yates*, que hoje é diretor de pesquisas do Yahoo na Espanha, no Yahoo Europa. Inclusive tem um livro, que foi mais vendido da área de *information retrieval* é justamente esse livro do *Berthier* com o *Ricardo*, o ‘Modern Information Retrieval’, que saiu uma edição nova agora em 2011. E o *Ricardo* é possivelmente, o top dois ou três mundial na área de *information retrieval*. E daí, tinha-se muitos contatos. Além do que, a gente era um grupo, somos ainda, um grupo muito forte academicamente. Então a gente participava de muitas conferências, enfim, participava de eventos. Acho que um dos principais artigos aí que saíram desse grupo, publicado na WWW de 2004, foi lá em Nova Iorque, um aluno de mestrado da *Gi Reis*, que já estava na coordenação nessa época, esse artigo chamou muita atenção do pessoal do Google. Tanto que o pessoal chamou o *Davi* para ir conversar lá no Google, em Nova Iorque, na época para mostrar o trabalho. Então acho que isso foi uma das coisas que chamaram atenção do pessoal, de que o grupo não era fechado. Tinha essa ambição. Era um grupo muito forte, muito forte mesmo.

C: Isso é muito interessante de saber e de ter essa relação. Eu vi uma entrevista do Alberto Laender na ACM, na associação internacional. Eu tenho alguns dados que mostram essa relação do grupo todo.

A: É, a gente continua trabalhando até hoje com o pessoal da UFMG. A gente mantém até hoje uma relação estreita com eles, nosso grupo aqui na UFAM, que é um spin-off do grupo de lá.

C: Gostaria de saber a opinião e as lembranças do momento que tem a ver com a minha pesquisa de história. Naquela época um dos mitos muito corrente era a questão do *ciberespaço*, se você for ver as matérias daquele momento vai constatar que tinha toda uma, vamos dizer, “mitologia” em torno do que seria esse *ciberespaço*, baseado em Gibson, no *Neuromancer* e outros autores. Eu gostaria de saber sobre essas ideias de que o *ciberespaço* iria democratizar, sendo quase uma revolução da democracia; elas faziam parte da vida de vocês na UFMG, ou aí na UFAM, como ideário? Você acha que vocês pensavam assim ou viam a tecnologia de forma diferente?

A: Bom Camila, eu acho assim. Isso nunca esteve, digamos, na frente não. Dizer que a gente tinha esse tipo de ideia como motivação para fazer a pesquisa, acho que não. O que a gente queria e sempre quis, e continua querendo, era manter o Brasil na vanguarda tecnológica nas áreas que a gente trabalha. Assim, o aspecto social disso, embora bastante importante, eu acho que não está nas mãos da gente. Está na mão de outro tipo de agente. Não dá para saber bem o

que será feito a partir daí. O que a gente sempre quis e essa era a ideia do *TodoBR*, de novo né, enfatizando... era manter uma vanguarda tecnológica na área. Para fazer uso para o país, o que fosse necessário. No fundo a internet de hoje, ela enfrenta o mesmo desafio, que é manter essa vanguarda tecnológica.

C: Agora faço uma pergunta geral, para o grupo: como pesquisador da área, que viveu também esse momento, nos primórdios da internet, como você lidava com essas ideias, por exemplo, de navegar, surfar na web, que era quase um *ciberespaço*, onde tudo poderia acontecer, onde alguns autores falavam na época? Hoje em dia este discurso já está datado, desta possibilidade de construir um mundo melhor, todo esse ideário. Como você se relacionava com isso? Pensando na sua vida e na sua experiência com a internet.

A: Eu sempre achei muito interessante a possibilidade da gente ter informação útil e disponível para todos. Então isso é um negócio que sempre me soou muito bem. Agora dizer que estava com isso, digamos na minha frente e quando eu motivei a fazer as coisas, acho que não dá para falar isso não.

C: Eu que agradeço, porque eu sei que é um esforço de rememoração.

A: É. Porque essas coisas assim, a gente segue em frente. O bonde anda.

C: E para finalizar, *Altigram*, te agradeço muitíssimo pela colaboração e gostaria de lhe perguntar duas coisas: o quê você acha que deveria ser preservado dessa história da web no Brasil? Códigos, páginas ou o quê você consideraria interessante? Fale também sua opinião sobre o estado da arte hoje nessa área de engenhos de buscas no Brasil, se você importante acrescentar algo.

A: Bom, eu acho que é super importante preservar a memória. Acho que, de novo como eu te falei, a gente não tinha muita essa idéia, mas olhando de uma outra perspectiva hoje, isso é super importante, porque no fundo é um pedaço da sociedade, cada vez mais relevante e que a gente não pode perder a perspectiva do que aconteceu. De jeito nenhum. Está fora de cogitação, a gente não ter esse tipo de memória, porque as pessoas tem que aprender com que foi feito no passado, o que funcionou e o que não funcionou. Então é importante que isso seja preservado sim. Da mesma maneira como que alguém algum dia pintou um quadro para mostrar como era a vida no Brasil colônia. Quero dizer que daqui 50 anos, como é que começou esse negócio de internet no país? É uma porção muito relevante da sociedade para você deixar ela relegada aos outros. Alguém precisa de cuidar disso, então parabéns pelo trabalho. Agora sobre a tecnologia de busca hoje no país, quero dizer, isso mudou muito. Mudou radicalmente. A *web* se transmutou em mil coisas diferentes. A rede social, a transformação da plataforma pro móvel, isso está causando uma revolução muito grande e aquele 45:09 [ruído- palavar não identificada], que

era a busca do passado, ele se desfazelou em um bilhão de fragmentos. Continua a pesquisa no *coRD* de busca, o *Edleno* que eu já menciou aqui, tem um monte de trabalho recentes nessa área. Existe toda uma comunidade internacional preocupada com isso. Aqui no Brasil, a UFMG e a UFAM estão na frente do processo. Mas muitas outras coisas relevantes estão acontecendo, por exemplo, todo o trabalho em torno de análise e de rede sociais, para a gente entender realmente esse fenômeno, como ele acontece, o que dá para fazer e o que não dá, para a gente saber discernir o que é verdade do que não é. O trabalho que as pessoas estão fazendo, por exemplo, nunca se imaginou que seria importante que é você monitorar a opinião das pessoas sobre as coisas, para aprender mais sobre elas. É um momento único, isso aí é muito especial. A gente está realmente vivendo uma época muito especial.

C: Te agradeço muitíssimo. Se não for incomodar e você tiver algo, um artigo, um contato, etc., que você ache que valha a pena me mandar, te agradeço por essa colaboração, nessa pequena aventura de contar essa história.

A: Parabéns para você também, porque não é fácil não. Ainda mais como é difícil extrair as coisas assim da gente.

C: Não é bem assim. Até então todos tem colaborado. Muito obrigada, tá!?

A: Abraços.

C: Abraços, até. Bom trabalho.

A: Tchau. Muito obrigada, para você também.

C: De nada.

Entrevista com Roberta Coelho, via Skype, em 17 de junho de 2014.

C: Boa tarde, hoje é 17 de junho e vamos iniciar a entrevista com Roberta Coelho, professora doutora de Ciência da Computação que vai contar um pouco da sua experiência na empresa Radix.

Roberta, muito obrigada pela colaboração neste projeto de doutorado. A primeira questão é aberta: você poderia contar sobre a sua experiência no Radix, a sua área de atuação específica, e quais são sua memórias daquela época.

R: De forma geral, foi muito enriquecedor trabalhar no Radix nessa época. E como foi praticamente o meu primeiro emprego, eu fiz meu relatório de estágio no Radix, eu não tinha ideia do momento em que eu estava vivendo. Depois, ao longo da vida é que fui perceber quão rico e especial foi aquele período no Radix. O Radix começou a partir de uma tese de doutorado, cujo o objetivo era indexar a web brasileira. Então ela começou como uma empresa em cubada no 1:14 [ruído], e ela tinha mais ou menos uns cinco funcionários, todos alunos de graduação ou recém formados. E nessa época era um projeto de pesquisa e inovação junto com o desenvolvimento. Então eu comecei a trabalhar na época que a empresa era bem pequena, no final de 99, que tinha já uns sete funcionários. Então a gente trabalhou com o desenvolvimento do Java e o objetivo desse projeto de pesquisa era gerar um produto que a se pudesse comercializar e fechar serviço depois. As tecnologias que a gente usou nessa época do Radix, realmente eram tecnologias que tinham sido compostas naquela época, as tecnologias web e a placa de recuperação de informação. Então, como eu terminei a minha graduação e já fui entrar nesse projeto de pesquisa, eu acabei desenvolvendo meu trabalho de mestrado dentro do Radix. Foi uma experiência muito boa e hoje, as pessoas como na empresa também, seguem bem as suas carreiras na Microsoft, Yahoo, outros ainda tem empregos nas áreas de desenvolvimento de software no Brasil. Realmente foi uma época muito especial, e depois quando eu sai de Recife e fui fazer Doutorado da PUC-RJ, depois fiz um sanduíche em Lancaster, e nessas viagens eu me dei conta do quão rico tinha sido aquele ambiente no Radix. Eu não encontrei nenhum ambiente igual aquele, nem no Rio, nem no exterior. Era um ambiente que você respirava tecnologia e inovação e ao mesmo tempo as pessoas se ajudavam muito. Eu acho que para a minha vida foi um momento muito especial.

C: Quem era o coordenador nesse período e dessa linha de pesquisa de doutorado? Fale um pouco sobre o seu mestrado. Você falou que tinha um projeto de doutorado que era o iniciador, o impulso; gostaria de saber quem era o responsável naquele momento.

R: Nós tínhamos um gerente de redes que era o Paulo Naves, ele era mais um gerente

de operações, como cronograma, delegar tarefas. E tinha um aluno de doutorado na época que também era professor, eu não lembro aonde, mas ele era professor em ciências 4:16, que era o Pedro Falcão. Então tudo começou com a tese de doutorado do Pedro Falcão que era orientado pelo professor Silvio Meira, da UFPE, que no caso acabou sendo meu orientador de mestrado. Porque a minha dissertação de mestrado foi diretamente relacionada a engenhos de busca. O gerente na época era então o Paulo Naves, o professor que tocava o doutorado e terminava sendo um gerente de mais alto nível, mais na parte técnica, era o Pedro Falcão. E tinha o Silvio Meira, que dava mais dicas estratégicas e reuniões motivacionais para falar para a equipe o quão importante era aquele momento, para contar com a colaboração de todos. E o que surgiu com esse projeto de pesquisa, depois foi vendido para a globo.com, para o site do IG. Todas essas transações comerciais, sempre era o Silvio Meira que repassava para gente as novidades e ele ficava mais nessa parte de negociações.

You are viewing an archived web page, collected at the request of [Universidade Federal do Estado do Rio de Janeiro](#) using [Archive-It](#). This page was captured on 10:10:10 Nov 28, 1999, and is part of the [Brazilian Web Search Engines \(1997-2013\)](#) collection. The information on this web page may be out of date. See [All versions](#) of this archived page.



Perguntas
Quais as buscas mais realizadas pelo RADIX?

Respostas
Os 100 campeões de "Hits" nas buscas do RADIX.

Categorias
Em breve. Aguarde!

Cadastro
Cadastre sua página no RADIX!

Agora com **4 milhões e meio** de páginas indexadas, o [RADIX](#) é o mais completo engenho de busca da Internet brasileira! Acompanhe o que está sendo publicado na mídia!

[RADIX](#) é tecnologia de classe mundial para encontrar informação na Internet, operando em fase de testes como uma unidade de negócios do Centro de Estudos e Sistemas Avançados do Recife, [CESAR](#).

Breve teremos a maior cobertura absoluta dos sites produzidos em língua portuguesa, com o índice mais atualizado, o processador de consultas mais preciso de menor tempo de resposta: tais objetivos conflitantes são resolvidos pelo conjunto de tecnologias proprietárias criadas pelo [RADIX](#).



Página arquivada pelo Internet Archive do Radix. 10/11/1999

C: Você poderia explicar resumidamente o que foi exatamente o seu projeto de mestrado?

R: Na época, em 1999, o engenho de busca mais popular do Brasil era o Cadê. O Cadê e o Yahoo, que era mais famoso do que o Google na época. Então, no Cadê e no Yahoo, na página inicial tinha um conjunto de links que se chamavam categorias, como esportes, educação e lazer. Na página inicial do engenho de busca do Cadê e do Yahoo tinham essas categorias que tentavam classificar os sites da web brasileira. O Radix na época, não tinha essa parte de categorias. Ou seja, só tinha caixa de texto, de fazer consultas. E aí eu entrei no Radix para implementar a primeira versão, do que a gente chamou de monitor de categorias. Seria permitir que a equipe de jornalistas contratados pelo Radix, que eles alimentassem essas categorias com sites interessantes. Então eu fiquei responsável por construir esse programa, que era alimentado

por jornalistas, que apareceriam na página principal do Radix. Essas categorias era uma parte interessante, mas ao longo do tempo a gente notou que mais importante do que as categorias era a qualidade da busca, a qualidade do resultado.

C: Existe tradução para esse termo em português? Eu estou usando “robô”, “agente”, “varredura”, “varredura automatizada”. Vocês usavam tradução para “crawler”, “capture”, ou “snapshots da web”?

R: A gente usava mais *crawler*, mas quando iria traduzir usava robô mesmo. Mas o termo *crawler* a gente usava no dia-a-dia. Tinha um aluno que ficava no *crawling*, outra aluna que ficava por conta da parte em que hoje no Google quando você digita uma palavra errada ai ele sugere uma palavra que você estaria querendo escrever e escreveu errado. Outra aluna fez nesse 9:53 [ruído: palavra não identificada]. Ai, foi que eu estudei na época o Google. Ele não era tão famoso quanto hoje e ele tinha uma interface bem limpa, era só uma caixa de texto que em que você fazia consulta, e para a sua surpresa os primeiros links eram aqueles que você estava querendo. Então a qualidade da resposta do Google, ele era muito além do Cadê e do Yahoo. A gente fazia a mesma consulta nos três e via que no Google aparecia o site mais famoso. Se você pesquisava sobre o clube Flamengo, o primeiro link era o do clube Flamengo. Eu ficava impressionada. O que que estava acontecendo? Como é que eles sabem disso? Foi ai que o Silvio Meira me deu um artigo do Sergey Brin e do Larry Page, que na época eram alunos de doutorado. O Google surgiu desse trabalho de doutorado do Sergey Brin e do Larry Page lá em Stanford. Ai ele me passou esse artigo e pediu para eu estudá-lo e ver como a gente podia melhorar a qualidade da busca do Radix, que era deficiente. Durante o meu mestrado eu estudei esse algoritmo do Google, e estudei outros algoritmos de ranqueamento. O meu mestrado foi eu desenvolver um algoritmo de raqueamento hibrido, que misturava algumas características do Google com características de outro algoritmo que era famoso na época que era o HITS. Então durante o meu mestrado eu criei e avaliei esse algoritmo usando o que a gente chama de Blind Test, que é o teste cego. Eu convidei os dois funcionários do próprio Radix e eles fazem uma consulta no sistema, sem saber qual algoritmo está sendo usado e ele classificavam aqueles links que retornavam em interessantes ou não. Que aí, a gente comparou a versão atual do Radix com a versão modificada, que era no caso com o algoritmo que eu desenvolvi, é o algoritmo que a gente chama de link analysis. Com a diferença do Google, a maioria dos algoritmos de ranqueamento só olhavam para características da formatação da página, ou seja, se naquele termo que você está procurando aparece em negrito, aparece um header, a sua página ganhava uma pontuação a mais e ela terminava sendo classificada no topo da lista. O algoritmo anterior do Radix só olhava características de formatação, usando uma lista. Se apareceu em negrito, ganha 10 pon-

tos. Se apareceu dentro de um link, ganharia isso. E o que a gente percebeu que o Google usava a estrutura de links da web com indicadores, para eles saberem o que era uma boa página. Ou seja, eles baixavam várias páginas. Você tinha uma página que era apontada por várias outras, significa que ela é mais importante do que uma que só era referenciada por dois ou três links. E no meu trabalho de mestrado eu fiz o algoritmo de análise de links. Ou seja, além de olhar a formatação da página, eu terminava olhando também se aquela página era apontada por muitas outras páginas, ou se ela apontava pra várias. Daí apareceu o conceito de página autoridade, que é aquela que é apontada por muitas; e página hub, que aquela que aponta para um monte de links. Então usando essas informações do *link analysis*, a gente conseguiu melhorar em 10% a qualidade da resposta.

INTERNET ARCHIVE
WaybackMachine

http://www.radix.com.br/busca/Categoria?pagina=1&catg=historia6.Co Go

NOV DEC JUN Close X
5
2000 2001 2003 Help ?

5 captures
5 Dec 01 - 14 Jan 03

Home | Coloque o Radix no seu site

Buscar

Brasil Mundo

Ajuda

Radix > Educação & Referência > Educação > Pesquisa escolar > HISTÓRIA

História do Brasil >>(34)

18 páginas (de 1 a 10)

Radix Recomenda

1. [História Net](#) História antiga e contemporânea, mapas e comentários sobre questões dos principais testes do país.
2. [Linha do Tempo](#) Da invenção da escrita ao fim do milênio, os eventos que marcaram a história do mundo.
3. [Net História](#) Notícias sobre história, arqueologia e paleontologia. Biografias, ensaios e textos.
4. [Vestibular Seriado](#) Traz um rotelro de estudo de história, mapas, questões de vestibulares e dicas de estudo.

Opção de caixa de busca e consulta ao diretório das categorias na página do Radix arquivada pelo Internet Archive em 5/12/2001.

C: Esse seu algoritmo ficou sendo utilizado pelo Radix? Passou a ser incorporado na metodologia da empresa?

R: Não, porque na época a gente tinha vendido para a Globo.com e para o IG. Esse daqui era um projeto de pesquisa FAP e cabia o gerente decidir se ele iria entrar ou não. Terminou que no fim esse algoritmo não entrou no sistema em produção, porque já tinha vendido. Pouco tempo depois, ou seja, menos de um ano depois, o Radix vendeu a solução e a empresa fechou.

C: Você tem alguma impressão desse momento? Hoje sabemos que o Google é responsável por mais de 90% das buscas no Brasil e no mundo, exceto em alguns lugares como na China

e na Índia, onde tem concorrentes pesados pela questão linguística e mesmo assim é bem forte. Nesse momento as buscas no Brasil naquele contexto o Cadê era o primeiro e foi vendido para o Yahoo, que também é uma empresa internacional. O Radix entra na medição do Ibope e da Mediametrics, mas ele não estava no topo. Fico me perguntando, como historiadora, como as pessoas que faziam parte percebiam ou não a importância de se ter mecanismos de busca nacionais.

R: A gente via a importância, mas sabíamos que nunca iríamos bater a qualidade do Google. Essa era a impressão que a gente tinha, por mais que a gente batalhasse, aquela qualidade do Google era uma coisa assim muito difícil de se...

C: Em 1999 e 2000 você já tinha essa impressão?

R: Já em 99. E outra coisa, eles não divulgavam o algoritmo. O artigo deles, era um artigo assim de idéias gerais, em questão de ver mais ou menos como era o algoritmo. Então a gente teve que ler umas três vezes para tentar decifrar as entrelinhas, que ele dava mais alguma dica do que estava fazendo. Desde o começo tinha muito segredo envolvido sobre como o Google conseguia aquela qualidade de resposta. E uma coisa interessante do Google era que ele era muito clean. Ele era apenas uma caixa de texto, e você fazia uma consulta. Já o Yahoo e o Cadê, que acabou sendo o modelo que o Radix seguiu, ele era uma interface muito poluída. Tinha muita informação na página principal e no meio daquelas informações é que tinha a caixa de texto. Isso a gente achava que poderia atrapalhar um pouco os usuários, mas desde aquela época a gente pensava como é que o Google consegue essa qualidade de resposta. A gente por mais que trabalhasse, quando fazia uma consulta no Google e uma consulta com um engenho de busca da web brasileira, mesmo em um engenho de busca focado para a web brasileira, ele não conseguia ser melhor que o engenho de busca genérico. E nessa época a gente não tinha muita ideia da importância desse período, porque o Google a gente via como um projeto de pesquisa, como qualquer outro interessante, mas a gente não imaginaria o rumo que o Google iria tomar. A gente não imaginaria, que hoje, ele seria essa empresa, desse tamanho, que não faz somente busca. É um monopólio. Então nessa época a gente constatava que o Google era muito bom, e por mais que a gente trabalhe vai ser difícil chegar a esse nível de qualidade, porque eles não divulgam tudo, e deve ter muita gente envolvida, muita coisa aí que não temos ideia de como eles fazem. A gente era aquele grupo de pessoas que saía no feeling, tentando buscar mais qualidade. O que aconteceu, na mesma época contemporânea ao Radix tinha o TodoBr, que era lá na UFMG.

C: Eu ia perguntar justamente isso. Vocês tinham contato entre esses buscadores, com o TodoBr?

R: A gente não tinha contato. A gente tinha mais assim, contato de saber que existia e era

tipo um concorrente do Radix. Mas a gente percebia que a qualidade do TodoBr não é superior a qualidade do Radix. Então era mais um projeto, e que depois eu fui perceber, que era um projeto mais acadêmico. O Radix virou mais empresa. O TodoBR, não sei se todas as pessoas que trabalhavam lá, trabalhavam na universidade. Porque no Radix, eu fiz meu mestrado, mas optei em não receber bolsa, e eu era funcionária do Radix, trabalhando 8 horas/dia. Eu implementava durante o meu trabalho, mas a parte de escrita e tudo mais eu fazia durante o final de semana ou quando chegava em casa e no final, pedia até redução na carga horária para ter mais tempo para escrever. A diferença entre o TodoBr e o Radix, é que o TodoBr era mais voltado para academia e não estavam preocupados com lucro. E já o Radix já saiu da universidade, alugou um prédio, tinha que pagar as contas, e questão do lucro, de vender e de ter clientes reais. Essa é a minha visão. E é tanto que depois a gente viu que o TodoBr foi comprado pelo Google. E o Radix no caso vendeu a solução, e o TodoBr, eu acho, vendeu o time que ele construiu e algumas ideias ou outras, mas nada que a qualidade dele fosse melhor que a qualidade do Google.

C: Não sei se você recorda, mas nesse momento houve alguma preocupação ou relação com a Universidade, por exemplo, com relação a essa indexação da web que você mencionou? Vocês teriam que fazer uma indexação preliminar para fazer buscas nessas bases de dados? Sobre esses dados que foram recuperados, naquele momento havia alguma intenção, preocupação ou questão de arquivamento para prosperidade?

R: Era mais a rotina o trabalho mesmo, que era fazer a indexação e armazenar as páginas. No começo a gente fazia o download das páginas. Mas essa parte da indexação, eu não sei em detalhes. Mas a gente não tinha essa preocupação. A nossa preocupação na época era os clientes, a Globo.com. A solução tem que funcionar lá no domínio da Globo.com e tinha o IG. A parte dos dados era mais assim, porque a gente tem que fazer porque o usuário vai fazer uma consulta, mas a gente não pensava os dados por si só, como sendo um bem da empresa. O bem principal da empresa na nossa visão na época, eram os crawlers, os algoritmos, e não os dados.

C: Você como professora desta área e pesquisadora, conhece alguma experiência no Brasil que trabalha dentro da área da computação? Alguma pesquisa sobre a utilização desde documento histórico, no nosso ponto de vista da sociedade? Um documento histórico que está em outro formato que tem que ser decodificado e trabalhado com uma técnica específica? Existem pesquisas sobre o arquivamento da web em diversas instituições, mas não é de meu conhecimento que exista no Brasil.

R: Eu também não conheço. E nem conhecia, porque meu mestrado foi em recuperação de informação, depois o Radix foi vendido, a empresa acabou. Ai eu terminei fazendo o meu doutorado em outra área, que foi na área de teste de software. Trabalhei como engenheira de

software e vi que os problemas dos softwares são os boxers. E acabei desenvolvendo o meu mestrado em outra área, não porque a outra área não fosse promissora, mas como não tinha mais o Radix e para ficar pesquisando nessa área teria que viajar para Minas Gerais ou Estados Unidos, sendo alguma coisa no Google. 26:37[ruído], na época que começaram mesmo o Radix, foi que ele fez o doutorado nele em recuperação de informação nos Estados Unidos. E até hoje, acho que até o ano passado ele ainda trabalhava e usava a mesma implementação de crawling adaptada, para fazer as pesquisas e escrever os artigos.

C: Como é o nome dele?

R: É Luciano Barbosa. No início mesmo do Radix, foram três alunos: o Oscar Miranda, que está na Microsoft; Luciano Barbosa e Tiago, mas o Tiago eu não lembro o sobrenome dele, e ele trabalha no Banco Central; e o Luciano trabalhou na TNT, no Yahoo, e hoje ele foi para o Rio trabalhar em um centro de pesquisa, não sei se é da IBM. E foram esses três alunos de graduação na época, que implementaram as primeiras linhas de códigos do Radix junto ao Pedro Falcão.

C: Qual é a sua impressão pessoal em relação a isso, que correspondeu justamente com o período de ascensão do Google no mundo inteiro? O Yahoo comprou o Cadê, o Google comprou o TodoBr, e o Radix desapareceu porque deixou de ser um mecanismo de busca; e, como já falamos, hoje o cenário é que o monopólio é do Google. Você vê alguma possibilidade de mudança?

R: A minha impressão pessoal, foi a de que eu gostaria de trabalhar no Google, porque eu acho que ele vai dominar o mundo. Porque assim, se naquela época a gente achava que já era uma coisa sobrenatural. Hoje em dia, então. Uma coisa interessante é que eu fiz parte graduação na UFCG, que é Campina Grande. E um dos professores da UFCG, que o Walfredo Cirne, que ele é essas pessoas fora da média, são essas crianças gênias, que se tornam professores e ele era professor da UFCG e trabalhava na parte de grids computacionais, *cloud computing*, nessa parte de distribuição de tarefas, essas tecnologias que envolve o que a gente chama de *server farm*, que são aquelas escalas cheias de servidores, executando processamento paralelo, de mil coisas ao mesmo tempo. Ele, professor da UFCG, tinha algumas publicações na área e de uma hora para outra ele recebeu um e-mail do Google, chamando para ele trabalhar no Google. Ai ele pediu um afastamento temporário da UFCG e foi trabalhar no Google com aquela ideia: Não, eu vou passar um tempo e depois eu volto. E ai, o tempo foi passando e passaram-se seis anos, que é o limite para 34:05 [ruído] e ele simplesmente não voltou mais. Então hoje, ele trabalha no Google com a família toda e ele é extramamente inteligente. Ele fez doutorado nos Estado Unidos, e a família dele recebeu até uma carta da universidade elogiando. Então eu pensei, se o Google tem um cara super inteligente como o Walfredo, imagina os outros. Então é muito difícil concorrer.

C: Hoje é claro que o nível de monopólio e de eficiência na busca é do Google, são dados indiscutíveis. Mas novas ferramentas começam a surgir, como o DuckDuckGo, um buscador que garante a privacidade do usuário; o slogan deles é ‘we do not track you’, enquanto o Google rastreia todo mundo, e afirmou desde o início que salvava todas as buscas dos usuários, e continua até hoje. Qualquer click.. Começam também a surgir buscadores ecológicos, que a cada busca que você faz, você doa. No Brasil eu desconheço essa pesquisa, mas nessa época o Radix rastreava o usuário? Como era a política em relação a isso do Radix?

R: Assim, a gente nem se preocupava com isso, com essa parte de privacidade na época. Eu não lembro se o pessoal de suporte armazenava o IP da máquina que é a sua identificação. Mais uma coisa que a gente sempre armazenava, eram as consultas que eram feitas. Porque nós queríamos saber qual era a palavra chave mais popular, quais são os termos mais buscados. Tenho certeza que a gente gravava as consultas.

C: Você faz ideia de onde possa estar essa base de dados?

R: Eu acho que se sumiu do mapa. Eu acho que foi formatada com os computadores.

C: Para mim, existem sempre os termos chaves de cada época. Seria uma ótima fonte histórica e muito interessante saber sobre quais termos a população fazia mais buscas em um determinado momento.

R: Na época do Radix, palavra que sempre foi popular era sexo.

C: Em todos os buscadores o termo mais buscado é “sexo”, mas também pesquisamos o número cinco e o número seis.

R: E é uma pena, porque a qualidade de informação de internet com tanta criança. E isso era uma coisa que a gente pensou na época foi essa parte de busca segura para crianças.

C: Vocês tinham isso no Radix? Hoje já existem buscadores com filtro.

R: A gente começou a implementar, mas eu não lembro em que pé ficou. Se a gente filtrava para todo mundo, eu não sei como ficou. Na época, pra gente os dados não eram importantes. O que se preserva até hoje é o código fonte. Para quem tava trabalhando lá no Radix, era muito mais importante o crawler, o código do crawler, do indexador, do que os dados, a história. E aí, os dados e a história ficaram no servidor lá e eu acho que foram deletados, ou eu não sei se as máquinas foram para a UFPE.

C: Sobre esses diretórios, gostaria de saber quem pensou na elaboração. Tinha o modelo em voga que era o do Yahoo, do Cadê, que eram esses sistemas de diretórios. Houve alguém responsável pela classificação? Você seguiram algum modelo?

R: Eu acho que o objetivo da gente era seguir o Cadê. O Silvio Meira teve a ideia do editor de categorias, e ele foi o meu orientador do TCC. Ele disse: Roberta, eu quero que você imple-

mente algo que permita que as pessoas criem categorias e subcategorias e adicione links. Na época tinha o Kiko, não sei se esse é o nome dele. Ele era tipo cliente do Radix, e que a gente passava as soluções e ele dava sugestões e tinha que passar pelo crivo dele. As iniciais mesmo, eu acho que foram baseadas no Cadê.

C: Eu sei histórias de outros buscadores, não necessariamente no Brasil, onde houve o trabalho de alguns bibliotecários. Vocês tiveram algum contato, por exemplo, com pessoas dessa área da universidade? Houve alguma preocupação em receber ajuda deles?

R: Não. Eu pessoalmente não tive contato, mas eu não sei se alguém foi procurar alguém dessa área. Eu lembro que existia uma preocupação com qual a melhor classificação. Eu nunca tive esse contato direto. E não sei dizer se alguém, como o Pedro Falcão, o Silvio Meira ou alguém da equipe de jornalismo entrou em contato. Porque na época a gente pensava que a equipe de jornalismo, o Silvio Meira no caso, era o grupo que saberia melhor inserir em conteúdo, selecionar informação.

C: Para finalizar, uma pergunta de praxe que faço para as pessoas que viveram, de dentro, os momentos onde os mecanismos de buscas começaram a ter maior importância social: no final da década de 90, os mecanismos de busca não eram vistos como algo tão importante. Gostaria de saber o que você lembra de pesquisar na web no começo, suas memórias pessoais e sua experiência; e o que você acha que deveria ser preservado da história da internet brasileira.

R: O que eu lembro de pesquisar na época, como eu fazia mestrado, eu pesquisava mais tipos de artigos e palavras chaves relacionadas a pesquisa, como *link analysis*. Como eu trabalhava e fazia mestrado ao mesmo tempo, eu não tinha muito tempo de usar a internet como forma de lazer. Sempre que eu olhava um engenho de busca, era mais um trabalho. Nunca ia pesquisar restaurante, como hoje em dia a gente pesquisa restaurantes no Google. Ou quer fazer uma compra de uma máquina fotográfica, pesquisa o ranking no Google. Antes então, o Google era voltado mais para pesquisa e mais uma forma de ter acesso a bibliografia, sem precisar de uma biblioteca. Que na parte de computação, todos os artigos são gratuitos, e então eu usava para baixar tais artigos e ler sobre os próprios algoritmos de busca na época. E o que eu acho que poderia ser preservado na web brasileira, mas também em toda a web. Bom, eu vou falar da minha impressão mesmo, seria a minha primeira homepage. Essas coisa que você mesmo não faz o backup, seria interessante se o Google fizesse o backup por você. Na minha curiosidade eu gostaria de saber como era a primeira versão da homepage da universidade.

C: Você sabe que isso se pode pesquisar. A da universidade eu acho até provável ter, porque os crawlers do internet archive eram voltados no começo para essas instituições maiores. A webpage pessoal é provável não ter, mas eu achei super interessante a sua resposta, e a partir

dela gostaria de fazer uma outra pergunta: quando você falou desse tempo da web, a unidade básica que veio para você é justamente a webpage, que, de fato, é a unidade básica do protocolo inicial da web. Hoje em dia e mesmo naquela época ao longo do final da década de 90, e sobretudo a partir do ano 2000, cada página é composta por elementos dinâmicos. Um dos problemas que vejo é como relacionar esse dados. Então, quando você imagina a web e como preservar a web? Nessa pesquisa de links, uma das questões fundamentais do L. Page era a relação dos dados, a relação entre os links. Sendo o link a unidade básica de análise. Que outras unidades básicas de análises você acha que seriam importantes? Que objetos digitais ou informacionais você acha que seria um elemento para preservação digital?

R: Eu acho que imagens. Elas seriam interessantes de preservar, porque os sites eram mais estáticos. Uma página 'linka' para outra e para outra. E com o passar do tempo a web foi mais sendo usada para abrigar não só os sites, como homepages, páginas de conteúdo estáticas. Hoje, praticamente muitos programas mudaram para os programas entre si serem executados na web, com isso tem essas tecnologias de *reload* parcial da página. Então você carrega só um campinho naquela página toda. E eu não consigo pensar em que eu gostaria de preservar em um site mais dinâmico, que seria tipo uma aplicação. Eu acho que a memória seria mais útil na preservação de sites estáticos, tipo Folha de São Paulo.

C: Acho importante ouvir as opiniões de pessoas que tenham uma convivência específica e especialista com o tema.

R: Eu uso pouco a web para executar uma aplicação. Eu, por exemplo, uso a versão web do Banco do Brasil, ou internet banking. Eu não ficaria chateada em perder a versão anterior, já que a atual tem menos BOBs, a interface é melhor.

C: Então para você a preservação do código ou das linhas do que foram criados internamente também não seria algo importante?

R: Não seriam. Para mim não. Pessoalmente, essa parte de aplicação para web eu não faria questão de guardar.

C: E para quem estuda ciência da computação, essa área de códigos, não seria interessante ter uma biblioteca de códigos?

R: Essa área da ciência da computação termina não sendo útil, ela se perde mesmo. Porque aquela tecnologia que foi útil dez anos atrás, você não vai ter uma razão para estudar ela hoje. Você sempre está querendo estudar a mais nova e sempre aquela correria que o conhecimento avança de uma forma que você não tem nem tempo de se manter atualizado. Então você terá menos tempo ainda para olhar o passado.

C: Entendi. Essa é a lógica das novas tecnologias. Muito interessante ouvir de alguém

de dentro também. Gostaria de agradecer a colaboração no projeto e pergunto com o quê você gostaria de finalizar essa conversa.

R: Gostaria de agradecer o convite e dizer que foi um prazer, porque para mim essa época foi muito legal, e eu acho interessante um trabalho de pesquisa que venha pesquisar um pouco sobre essa história, esse momento e trazer para análise esse período, onde muitas coisas aconteceram. E para quem estava vivendo na época não tinha ideia que hoje a vida das pessoas estariam associadas a qualidade do engenho de busca e ao impacto que os engenhos de busca terminaram tendo no dia-a-dia.

Entrevista com Luciano Barbosa, em 25 de junho de 2014.

C: Bom dia Luciano. Hoje é dia 25 de Junho de 2014. E eu vou entrevistar Luciano Barbosa, sobre a sua experiência no Radix, uma ferramenta de busca criada em Recife. Primeiro lugar, Luciano, obrigada por participar dessa pesquisa. Eu vou começar pedindo para você se apresentar, dizer a sua formação e o ano que você entrou no Radix, a sua função, essas informações mais básicas.

L: Minha formação em Computação, eu fiz graduação na UFPE em 96 até 2000. E em 2001 a 2003 fiz mestrado lá também, e em 2005 a 2009 eu fiz doutorado na Universidade de Utah, nos Estados Unidos; isso tudo em Computação. Em relação ao Radix, eu posso falar que eu fui um dos que criou o Radix. O Radix inicialmente era um projeto, chamado projeto BRight, que era de um estudante de doutorado, chamado Pedro Falcão. Ele é estudante de Sílvia Meira e da Ana Carolina Salgado, professores da UFPE. Então nesse projeto BRight eu fui o primeiro, vamos dizer assim fora o Pedro, o primeiro estudante a entrar, através da iniciação científica.

C: Você foi bolsista de iniciação científica do CNPq na graduação?

L: Isso, e foi em 97. O projeto BRight foi o projeto que iniciou o Radix. Em 97 ele fez esse projeto e depois entraram mais dois colegas meus, Tiago Vieira e Oscar Miranda. E nós três fomos os que implementaram o primeiro núcleo do Radix.

C: E a sua função nesse núcleo inicial estava relacionada a que área?

L: O projeto de doutorado do Pedro Falcão era na parte de sistemas de busca distribuídos e eu fazia parte da implementação desse sistema.

C: Quais são suas principais memórias desse momento inicial do Radix? Como que vocês percebiam a construção de um sistema de busca distribuído, local? E como era sua relação com o que estava acontecendo no resto do mundo, assim o desenvolvimento já em 97-98, com o Google?

L: Naquela época não tinha o Google ainda e o engenho de busca mais famoso fora do Brasil era o Altavista. No Brasil tinha o Cadê, mas o Cadê não era bem um engenho de busca.

C: Mas 98 já tinha o Google, não?

L: 97, quando começamos, não tinha o Google. Tinha o Altavista, que era o mais popular. O Google só foi ficar popular por volta de 2000.

C: Tinha o Lycos, o Yahoo.

L: Tinha o Webcrawler.

C: E no Brasil era o Cadê? E estava sendo desenvolvido na UFMG, o TodoBR.

L: Isso. A diferença era de que o Cadê era comercial e o TodoBR e o Radix eram mais acadêmicos. O TodoBR foi muito agraciado com dinheiro do governo e o Radix recebia do

Opportunity, por exemplo. E foi o Opportunity que bancou o Radix, quando saiu do BRight.

C: E nessa época, da construção inicial, vocês tinham contato com outras pessoas de tecnologia da informação no Brasil, como as pessoas do TodoBR, ou o pessoal do Cadê? Como era a relação?

L: A nossa relação era mais acadêmica. Víamos o que eles faziam através dos papers, mas não tínhamos contato algum.

C: E com organizações internacionais, outras universidades fora do país, ou outros grupos? Tinha alguma relação que se estabeleceu nesse momento inicial?

L: Não, não tinha. A única coisa que teve foi que o Pedro Falcão foi fazer doutorado sanduíche em Columbia e lá ele fez alguma coisa. Essa foi a única relação com uma universidade de fora.

C: Hoje, nos já temos claro a importância de mecanismos de busca para a sociedade, e do Google. Naquele momento, que era um momento de pioneirismo, de desenvolvimento dessas tecnologias, como é que vocês viam a função de um mecanismo de busca, que nos papers eram chamados de engenhos de busca? E como essa nomenclatura de tradução foi feita? Como vocês lidavam com isso?

L: A nossa maior motivação na época para criar um engenho de busca, aqui no Brasil, era porque os engenhos de busca de fora não cobriam bem a web brasileira. Então a motivação do Radix e do TodoBR era ter uma cobertura boa da web brasileira, que na época tinha umas milhões de páginas. Não era uma coisa tão grande, mas como os engenhos de fora estavam preocupados em abraçar o mundo, eles não conseguiam ter uma cobertura tão boa da web no Brasil. E uma coisa boa disso, é que o Radix estendeu isso. Nós estendemos para notícia, fomos os primeiros a criar busca para notícias, busca em Português. Tínhamos algumas especializações no engenho de busca para alguns nichos, para música, para mp3.

C: O que era super novo para a época. Uma inovação dessa tecnologia. E desse momento inicial do Radix, o que que você lembra do desenvolvimento dessas tecnologias, comparado com o que que estava acontecendo lá fora? Vocês acham que vocês estavam correndo paralelamente, ou houve um gap com o Google? Como você vê essa situação depois de crescer o domínio do Google?

L: Problema de comparar com os outros é que a gente não sabe o que acontece por debaixo da cena. Então a gente só sabe quando tem um tipo de release e o pessoal comenta. É difícil comparar, mas a gente só podia comparar com o que a gente tinha, como a funcionalidade que tínhamos. Funcionalidade em busca de notícias, em música, etc. A gente via que era um diferencial, que ninguém tinha na época.

radix
A Internet na sua Língua

Procure Por: em: o que:

As 100 perguntas mais frequentes

O Radix mostra as páginas, mp3 e imagens mais procuradas pelos usuários.

Você pode conferir as buscas feitas ontem ou o total de pesquisas feitas na semana e no mês clicando nos links abaixo:

O que: [páginas](#), [mp3](#), [imagens](#).
Quando: [ontem](#), [na semana](#), [no mês](#).

Páginas: [1](#) - [2](#) - [3](#) - [4](#) - [5](#)

	Palavra	Consultas
1	receita federal	12841
2	www receita federal gov br	5399
3	cpf	3150
4	receita	2458
5	recadastramento cpf	2098
6	receita federal cpf	2082
7	www receita fazenda gov br	2028
8	receitafederal	1719
9	receita fazenda	1366
10	site receita federal	1230
11	imposto de renda	978
12	receita federal gov br	729
13	recadastrar cpf	669
14	secretaria receita federal	643
15	recadastramento de cpf	590
16	secretaria da receita federal	578
17	recadastramento	522
18	playboy	479
19	recadastramento do cpf	440
20	fazenda	430

Natal no Radix!

Cultura
Os melhores links e novidades sobre cinema, música e outras artes.

Sala de Aula
Encontre a resposta que o professor pediu para amanhã.

[www.jornalink.com.br](#)

no.com.br
no.com.br, n de notícia e o de opinião. Informação, humor e inteligência na Internet.

Colunistas:
[Marcos Sá Corrêa](#)
[Ancelmo Gois](#)
[Tutty Vasquez](#)
[Villas-Bôas Corrêa](#)
[Arthur Dapieve](#)

Página arquivada no Internet Archive, 02/12/2000.

C: E a sua pesquisa de mestrado foi exatamente sobre o quê?

L: No engenho de busca, uma preocupação é fazer a atualização do índice das páginas. Existem várias técnicas para se fazer isso, e a minha dissertação de mestrado foi sobre uma técnica, usando algoritmos distribuídos, para poder melhorar a qualidade da atualização das páginas de engenhos de busca.

C: E essas bases de dados que vocês iam compondo ao longo do tempo, o que acontecia com essas bases de dados depois que elas eram atualizadas? Esses dados anteriores eram automaticamente deletados?

L: Exatamente. A gente não tinha espaço para ficar guardando histórico.

C: E você tinham alguma relação com algum outro ramo de pesquisa da universidade que não fosse de pesquisa ligado a Ciência da Computação? Por exemplo, área de Humanas, ou pesquisa social, ou de estudos da Ciência da Informação, ou Bibliotecas?

L: Não. O único pessoal fora da área da computação que a gente tinha contato e que participou do Radix foram os jornalistas, responsáveis pelo conteúdo. Fora isso, a gente ficou muito isolado com o pessoal da computação mesmo.

C: E no Brasil, o que aconteceu com essas bases de dados? Você sabe de alguma coisa?

L: Não sei. A única coisa que eu tenho do Radix é o código.

C: Mais nada além disso você acha que sobreviveu?

L: Dados não. Uma coisa que talvez você possa pesquisar, não em relação aos dados, é que o Radix fez propaganda na Veja e anúncio na televisão.

C: Eu até que consegui alguns artigos de jornal do Radix e algumas páginas no Internet Archive. O Internet Archive salvou alguns snapshots, mas eles são muito incompletos. Uma das minhas curiosidades era saber como foi feita essa indexação das categorias? Como era feita essas categorias? E depois quando a busca passou a ser feita pelo sistema dos algoritmos, o que que modificou? Você tem alguma ideia, se houve muita mudança ou pouca?

L: Eu realmente não lembro disso aí. Acho que a Roberta Coelho que trabalhava mais com essa parte das categorias.

C: Você trabalhava também nessa parte da arquitetura de busca ou não?

L: A minha praia era mais a parte da indexação, do crawling. Porque tem dois componentes básicos, na verdade são três. Tem a parte do crawler, que é o programa que vai pegar as páginas na web; uma vez que você tem as páginas você vai criar o índice e o Oscar era responsável que hoje está na Microsoft; e tinha a parte de busca. Nessa parte tinha várias pessoas, o Marcelo Fernandes e que hoje está em Natal. Tiago participava comigo na parte do crawler.

C: Sobre essa parte específica do crawling. Você programava para que tipo de páginas o Radix? Vocês privilegiavam, por exemplo, o domínio “.br”? Ou a partir do in-links? Porque as páginas do domínio “.br” tinham autolinks para as páginas da web em geral. Como vocês programavam esse crawling?

L: A primeira coisa é que a gente restringia ao domínio “.br”. A gente não pegava página fora desse domínio, isso para o site principal. Para o site de notícias a gente já pegava o domínio “.com”. E a gente tinha o objetivo de deixar o crawler rodando, então boa parte do nosso trabalho foi verificar se o crawler estava rodando, se o crawler estava baixando as páginas corretamente, deixar ele rodando 24/7. A velocidade também era muito importante para a gente,

então a gente mexia muito com a parte de performance do crawler, mexia no código para deixa ele mais rápido.

C: No retorno desses resultados, para o usuário, nesses algoritmos, as variáveis que entravam no caso do Radix, como que entrava o peso da publicidade, ou da escolha, da seletiva dessa equipe de jornalistas? Qual era o peso desses fatores?

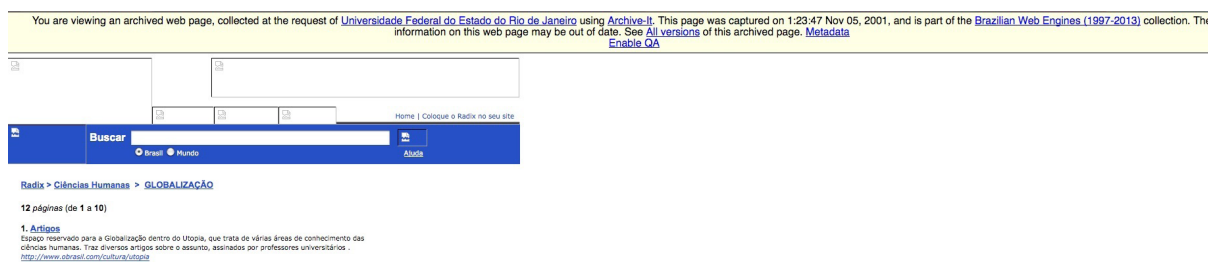
L: Eu lembro muito bem disso. Uma coisa que eu lembro é que categorias tinha um peso muito grande no resultado, na eficiência, no ranking. A gente não usava links, a gente usava basicamente a relação do texto e informação com categorias.

C: No Radix havia propagandas em cima dos resultados?

L: Sim. Já tinha vários banners.

C: Entrava como banner ou como resultado?

L: Entrava como banner.



Página arquivada pelo Internet Archive em 5 de novembro de 2001.

C: Isso é difícil de ver nos snapshots do Internet Archive. E sobre a sua pesquisa de doutorado. Qual foi sua área? Como a experiência no Radix pode ter influenciado essa pesquisa?

L: Então eu posso falar um pouco sobre o fim do Radix que está relacionado com o meu doutorado. Em 97 começou o BRight que iniciou o Radix, em 99 nós nos tornamos Radix e fomos para o C.E.S.A.R na UFPE, e em 2000 a gente foi para Caxangá. Tem um cientista que trabalhou com a gente lá, o João Batista e hoje é professor em Feira de Santana, que disse assim: “O radix começou no dia da mentira, 01 de Abril de 99, e acabou no dia das bruxas, 30 de outubro de 2002”. Então em 2002 eu estava no final do meu mestrado, o Sílvio Meira que era presidente na época do Radix, chamou a gente para o auditório, e não estava nesse dia, e simplesmente estava todo mundo demitido.

C: Nessa passagem, não houve uma perspectiva de manter a empresa depois da venda?

L: Foi um processo contínuo, foi acontecendo. O Radix foi vendido para uma empresa em Santa Catarina que eu não lembro o nome, uma empresa grande de softwares.

C: E o que aconteceu com a tecnologia do Radix? Você tinha o código que era dessa época, e depois que a empresa acabou, o que aconteceu com essa tecnologia desenvolvida?

L: Eu ainda dei consultoria, porque a gente tinha que instalar ainda na Globo.com. Criou-se uma empresa com ex-funcionários, até me chamaram, mas eu não quis e não estava interessado.

C: Você sabe o nome dessa empresa?

L: Eu não lembro. Eu sei que essa empresa ficou com o legado do código e dos contratos do Radix. Continuaram tocando e depois de um ano eles fecharam.

C: Como uma empresa que tinha uma tecnologia de ponta e inovadora, de repente desaparece, dissolve?

L: Como o nosso foco era o mercado brasileiro, e o dinheiro que a gente ganhava era de investimento, então conta-se que o Radix acabou porque o dinheiro que vinha do Opportunity, depois do 9/11, dinheiro para área de risco sumiu no mundo.

C: Estourou a bolha da internet.

L: Exatamente. Então depois disso já não tinha renda e nem mercado no Brasil para sustentar a gente na época. Até mesmo com o TodoBR, eles só continuaram porque eles tinham muito dinheiro do governo, e bolsas, muito mais acadêmico mesmo; e depois o Google comprou.

C: É engraçado, porque é um mercado que estava crescendo muito no momento, esse mercado da internet no Brasil estava em crescimento. Mas de fato a questão de investimento era precária.

L: É como conseguir dinheiro para se sustentar. O Google é, basicamente, uma empresa de *advertisement*.

C: É, mas o Radix e o Cadê tinham publicidade.

L: Isso, mas não se compara ao Google. E o mercado brasileiro era ainda pequeno, não tinha muito cliente. Não sei como é hoje.

C: Hoje, o Google.br domina 91% do mercado. É praticamente um monopólio.

L: A minha dúvida era mais em relação a captação de dinheiro. Se o Google.br se sustenta com o próprio dinheiro de advertisement aqui no Brasil. Ou se ele precisa de dinheiro dos Estados Unidos.

C: Esse é um ponto... isso eu não sei responder.

L: Então voltando naquele ponto lá, o Radix acabou em outubro de 2002, e eu estava no fim do meu mestrado. Corri para terminar meus experimentos antes que fechassem as portas, porque algumas coisas eu rodava na máquina de lá. E aí, a Ana Carolina Salgado que era uma das donas do Radix, ela era a minha orientadora de mestrado. Quando eu terminei o meu mestrado em Março de 2003, eu falei para ela que eu gostaria de fazer um doutorado fora do Brasil.

Eu cheguei a chegar a procurar sozinha e mandar email para outros professores, mas tinha uma professora brasileira lá nos Estados Unidos, chamada Juliana Frei. Ela na época tinha um aluno que fazia um doutorado sanduíche com ela, o Denilson Barbosa que é hoje professor na Universidade de Alberta. Então Carol mandou um email para o Denilson, e Juliana estava procurando alguém para trabalhar com ela em doutorado. Ai, foi perfeito.

C: E seu doutorado foi em qual área?

L: Foi na área de deep web. Que são os conteúdos que estão por trás dos fóruns. Meu doutorado foi basicamente sobre isso, sobre como fazer crawling desse tipo de conteúdo, como você acha os fóruns, que são a entrada para esse tipo de conteúdo, como você analisa e etc.

C: E você utiliza essa pesquisa hoje, no Brasil?

L: Eu fiz meu doutorado lá em Utah, nessa área de deep web, e em 2009 eu fui para a AT&T, no laboratório em Nova Jersey. E lá eu usei algumas coisas, mais da parte de crawling que eu fiz no meu doutorado.

C: E nessa época, você já se desligou um pouco do que estava acontecendo nessa cena de mecanismo de busca no Brasil. Depois disso, você soube de alguma outra iniciativa, do Brasil, no sentido de criar um mecanismo de busca, ou nessa área de pesquisa?

L: No Brasil, basicamente, esses engenhos de busca saíram de centro acadêmicos. O Radix da UFPE, o TodoBR saiu da UFMG, que é o melhor departamento de computação no Brasil. E o pessoal de Manaus que são filhos do pessoal da UFMG, como o Altrigram Soares, o Edleno Silva, eles ainda fazem alguma coisa nessa área. Se tiver alguma iniciativa, será ou do pessoal de Minas ou do pessoal do Amazonas. Tanto porque o Google. br em Belo Horizonte, não é tão ligado ao pessoal da UFMG. Então se surgir alguma iniciativa ou sairá de dentro da UFMG ou da UFAM.

C: E nessa época que vocês estavam trabalhando com a internet, que leituras fora da sua área técnica você lembra que eram importantes para pensar o que eram crawlers, a deep web? Algum livro importante que você leu de ficção científica ou um filme, que chegou mais aqui no Brasil?

L: Coisa externas?

C: O que você lembrar...

L: Eu, realmente, não tenho muita lembrança disso não.

C: Nessa época tinha algum fórum de discussão, ou uma revista que você achava interessante pensar sobre a internet? Porque essa época tinha muito aquela crença de que a internet ia salvar a humanidade através da democracia, democratização da informação.

L: Como a gente estava no dia-a-dia, tentando matar um boi ou leão por dia, a gente não pensava nessas coisas. A gente estava muito focado naquilo ali. Profissionalmente, talvez tenha

sido a melhor época da minha vida. Mas era muito stress. A gente estava mais preocupado com os problemas do dia-a-dia. O Sílvio Meira com certeza estava pensando nesse tipo de coisa, porque ele não estava no dia-a-dia.

C: Você gostaria de falar mais alguma coisa? Sobre esse período, sobre os primórdios da internet no Brasil. O que você acha e lembra, e gostaria que fosse preservado? Você tem alguma nostalgia de navegar naquela web antes da web de hoje? O que você pensa sobre isso?

L: Nostalgia eu não tenho não. Acho que a web mudou para melhor, tem muito mais coisa. Eu tenho nostalgia da amizade, do grupo lá. E a minha lembrança inicial da internet foi em 1993 que era nem internet direito aqui no Brasil, meu primo tinha acesso ao BBS. Entrei na USP em 94, que eu não tinha acesso a internet. Só fui ver internet em 96 quando eu entrei para a UFPE.

C: Te agradeço muitíssimo, Luciano. E se você quiser falar uma última coisa, está aberto.

L: Gostaria de agradecer. Foi um prazer falar sobre esse período, que foi um período muito feliz da minha vida. Boa sorte e se precisar de mais alguma coisa, contato, pode pedir.

C: Super obrigada.

L: Obrigada você.

Entrevista com Tjarda de Haan, no Museu de Amsterdam, em 28 de fevereiro de 2014.

C: It's 28th February 2014 and I am going to interview Tjarda de Haan, coordinator of the Webarcheology project of the Amsterdam Museum. Thank you very much for collaborating on this research project ...

E: Very welcome and thank you for asking to give a presentation about the Internet pioneering subject within history, like the *Digital City*. And the digging up of the *Digital City*.

C: Thanks.

E: This [showing imagens on her computer screen] is the presentation I've been giving in Riga, Latvia, because there was the media art history conference and we were invited to talk and you have short talk... but at least it's concise.



Website do Museu de Amsterdam. Disponível em: <http://hart.amsterdammuseum.nl/71616/nl/bewaar-als>

E: The history of the *freeze* will be within this presentation and always we brag a little about how great it was used and how forward in the *e-culture* the machine was operating, like on the web next to the corporate site we have this co-creation platform, probably you've been there, with all the color faces, those are the avatars, we dug up within the process, socially we strong mobile, we have apps so you walk through the city and you get the history and we're building a *semantic-web*.

C: The mobile is a plan for...?

E: It's an app. We have different apps.

C: Already?

E: Yes, already... where we have integrated our collection, as well as when you going to have walks in Amsterdam and you have the data with you.

C: That is great.

E: So, we try to be where... you'll be surprised with what internet can bring us. And with this project mostly I've started with some pictures and asked the audience like do you recognize these pictures? Because I bring people back to 94 and that was the time that there was this telephone company., which it is not there anymore or at least has a different name and has reorganized itself, and you have to put on the telephone lines on the ground, and here there is an image, mostly most people had never seen it and don't recognize it. And it is some person sitting on a chair, which I am going to show you later on because that is one of the results of the project. We have integrated the *Digital City* already in the collection, the physical collection of the Museum. This chair, which you will see later in the presentation, this was called *the public terminal*, because in 94 if you did not have a computer at home or any modem or technical ... you do understand, you could go public and they have put six of these chairs with a computer in the city center, so you could go there. And they were very popular by gamers and it was a line of people waiting to get that chair.



Disponível em: <http://hart.amsterdammuseum.nl/71616/nl/bewaar-als>

C: And this chair was put by the *DDS* or by the municipality? Or it was a project together with the government of Amsterdam?

E: Yes, I will explain later on, but it was started with media people of the *De Balie*, a cultural debating center, and the people of *Hack-Tic*, which was a hacker's magazine. And they got money from the government. But they started by a fluid group, there were a lot of people involved but *Marleen Stikker* was the initiator and was given the assignment by *De Balie* to think about how the electronic debating can influence here, and just figure out what is going on in their own states, with the ability of more systems and debating on the intimate.

E: So, this was *the public terminal* with the computer, so this is an image and it says so much. Because there is a person, it's a homeless or a person who tries to disguise its face for certain reasons. So I don't have the star yet, but we are interviewing and hoping to find out the story of this specific person of this particular photo. And this was the first interface of the *Digital City*. And there was no *world wide web* as we know it now, yet. So this was super important system based on the *MS-Free.Net* from Canada and States and, here the hackers escaped usually because the hackers already knew how to build up the *BBS*.

E: And this is a picture... we were in the international audience, just to show the domain names in The Netherlands. We have the numbers from 96, because then the organization started officially to noting it. But of course there will be four. But you can see, you could finish reading the intimate at that time in Holland, we have not even ten thousand websites. But website thinking, because this is a project of the *World Wide Web*, because it is about the digital object. So, you see we are in the dark ages of doing research, and this is my introduction way back in *web wide cyberspace*, there was city from network cables, it's like who've done it. Because we are doing forensics in a way, *digital forensics*. Trying to find and reconstruct the City. Very important and one of the reasons in this project is help our digital heritage. In 2003, UNESCO wrote a chapter, you know, I don't need to tell you...

C: The Charter for the Preservation of Digital Heritage ...

E: Yes, so, I summed up what are the difficulties. Well, the material, it is weird and complex. And the methods, there are certain methods developed and you asked me in the e-mail about *Brewster Kahle*. I've met him and I've talked about our project and probably I'll meet him in May. Are you here in May?

C: Yes.

E: Because there will be the next web conference and he will be one of the key speakers.

And I'm trying to set up a meeting with him. Later, we will talk about it. So, what he is doing is..millions of pages in his website. He is making *snapshots* and what we call that is the *web archiving*. And what we want to do, what I just did, is to reconstruct the whole City. Completely different thing... 8:26 [ruído]

C: Can I ask a question? I wonder about the Dutch policies for digital heritage.

E: I don't know the name of the organization...

C: No problem.

E: We have *Digital Erfgoed Nederland*, based in Den Haag. I should know their story better, but I'm not sure even when did they were grounded, but... early days, in 2010, there was an official piece written in assignment of *Digital Erfgoed Nederland*, and they presented the report, and as well like: watch out, we should act now! So, I can send you those reports. 09:53

C: You were working...

E: That is so good about Amsterdam, there are a lot of organizations, but also some cultural organizations involved in cultural heritage, but we have a little major setback with the cultural department. [ruído] So, the methods... well, we are trying to recreate a new method and the efficient task, who is doing what...because if you are talking about digital heritage, who has the software library...because we are dealing with software from years ago...who has it? And it is not affordable for private person to archive, because you have to pay licenses, updates...so that would be suitable for a library to do. So, we are hoping for a software library that will start and as well the hardware. About the freeze, just to get a little in the freeze, we did five free events, with beer...We were so happy. In one week we were so happy and one week later we started to work again, and we were like huh. Where is the tape reader? How can you play a tape from 69, sorry 96.



Website do Museu de Amsterdam. Disponível em: <http://hart.amsterdammuseum.nl/71616/nl/bewaar-als>

T: So, with a lot of partners, like the *Royal Library*, in Den Hague, we have a very good connection, and with *UVA Computer Centrum*, we have a good connection too, so in a way... but I am going to tell you about it later, because that is the history of *the freeze*. So, who is doing what... we need a *hardware*, library as well and *UVA, the University of Amsterdam*, has a computer museum... and, *Stichting Computer Erfgoed Nederland*, they have register online, where people can put their information, such as: what kind of computer do they have at home, what kind of hardware do they have at home... So you can, at least, know where some piece is, but that is also still being on the construction... how to make sense of a lot of hardware and where is it.

C: What is your methodological approach?

E: Approach. There is no approach. We just try and try, error and error, and we've seen it a lot. So *Digital City*... this tells me, I know this already. *The Digital City* first public virtual city in the world and the first Dutch online community and it gave access to the internet, to a big audience, because Universities were there, the Government has access for the defense, so there were certain people that had already access because of the perfection, but because the big audience did not have access. So, *DDS* with extra soul they gave this access to people. So it was like... for a lot of people the first experience online... Grounded by a fluid group, *Marleen Stikker* was the initiator and a lot of people she has spoken to and were helping and doing their own things as well.

E: CNN. We have a very nice video online and CNN was in The Netherlands in 96/97

and they made this comparison that the Dutch has helped the international trade and the way to opened up, and now we are doing that exactly the same for the internet. It is a really nice video! Then *Manuel Castells*, did the famous scholar that I was telling you and he has also written about the *Digital City*, like being across of the movement and that was very specific of the *Digital City*. And *inhabitants*, you know, almost as much as Facebook has now... it's incomparable, but a hundred forty was a lot for their start.

C: It is amazing.

E: You're amazed?

C: yes.

E: Here [showing scanned image of a piece of paper]This was the telephone number you should dial... of course, but this is a famous telephone number and here... I'm just going to mention the *freeze* in between, you know... because the *freeze* was made in 96, and it was because there was celebration of two years of the *Digital City*... and they thought like, we've been working a lot the last two years and we had three major changes in the City, like there has been three interfaces, not only interface but also from the *BBS* to the second interface and then to the most famous interface, this is the *sitemap*, this is not on the front page, but this is the *Digital City 3.0*... And this was built on squares, because this is also more getting the felling of the city, well but we don't really have a lot choice in our interface planning, because we have numbers, and you go... I am not sure with you are familiar with the *BBS*, but it had many ways when you could go, of course the help desk with you were panicking, you have the library with all the...

C: So, it is like you imagine like how that interface could work, but it is like I never navigated on...

E: We hope, but that is about the reach and the money we can find for the project, and if we can also reconstruct this, but mainly our focus is to reconstruct the *Digital City 3.0*... Because that is...

C: more interesting in a way...

E: both are very interesting and I would really love to have the two next to each other but here we knew we can find the data from this, and the original data... so, we hoped we could find it. So, we go to the *Digital City*. And that is complex enough, you know... It is a lot of work...

C: I can imagine.

E: And this was actually the first interface... Ow, that is not a good picture; it is on the website I think... This was designed by *Wout van der Kruis* in 93, but was no *mosaic*, was no possibility to make a user crack the interface... so, of course we started in 94 with the *bulletin board system*, two months later they have *mosaic* and they directly followed the technological

changes and implemented... and they could now go more to the experience of the City... and that was very important... 17:14 [ruído; nome não identificado], she was the designer; she made a very nice city... I can show you maybe later...

C: Please.

E: This is the *plattegrond*, so this the site map. You had a café, where you could chat, meet, discussion over the news... and some strange topics... where you could go... and you could see who was here, people had their mail. So a lot of people also used the *Digital City* just for their mail, but you can also build a house. And a house was actually a website.

C: And, how much of this was “saved” by by the Freeze?

E: So the *freeze*, they were in 96, and we have to make *still*. And a *freeze* is like *software still*, right?! Two years ago that they started and they had already three interfaces. And we don't know how many interfaces will come, so they decided we have to make a *still* for the future archaeologist, so they know what it look like. And that is very nice, because we are the future archaeologist now. So, it is a very nice turn.

C: When you are talking... I know you had the chance to talk with some of these pioneers... do they tell anything about this desire to conscientiously build a memory for the future... because I was curious and the internet archive was also created in 1996... and I was also researching like the term, *born-digital*... and I went to the N-Gram and it is also in 1996... so, I was curious to see how this term appeared at that moment... in different places, and about this initiative.

E: Which term?

C: *Born-digital*

E: Like, how it started?

C: If you research, for instance, in Google N-Gram, the database of books and news for the XX century, if you put the term we see that the first time it appears in press that was also in 1996.

E: ok.

I was also curious if they have a term in Dutch...

E: No, we follow many English words like computer.... 19:49 [ruído] *Born-digital*.... in Dutch? No, never.

C: That is good to hear...

E: If you find...

C: I'll tell you all about this!

E: So, just for showing you the change in history of the City interfaces. And here you have more of the digital frequency, and here you see the paper prototype. The photo is bad, it is better online. How they started to build the City? I still don't have the name of the photographer, but we also want to find it. Here you have many squares, the national elections, the sport square, the Central with a lot of 20:58 [ruído] and also the death square. We are interviewing the...

C: Sorry, what is the death square?

E: The place where you put memorizing bouquets for the people and all news about death... things... We are interviewing the designer, how he is started with design. Did he get an assignment...and how did he start it and how he was inspired. And that we will put online as well.

E: So, the goals of the project... I think probably you read this already on the website. Like, we preserved the *internet-historical monument*... and then we are so proud! I will show you like, ow!

C: I really want to see it!

E: And to map the history of the *Digital City*, internet and *e-culture*. Because the *Digital City* is very unique. It is a very clear case, but then Dutch. It is Amsterdam history heritage, but at the same time it is telling the history of internet in The Netherlands. How did it start? With

whom? How did it evolve? And *e-culture* is a great part of the Museum. And we are telling that story as well. And then, also included in the collections other heritage, because the Museum has its own specialties, and can preserve some of the collection. But we are not data experts, like the *Royal Library*. Or the *Stadsarchief* in Amsterdam; and you have the *IISG*, which is the International Institute of Social History. So, there're partners in the project. Well, we work together... and they told us: please, if you have the collection, with you have the City, if you reconstructed the City, please as well, put it in our collection. So, that is good for us, because they will safeguard it. They have the experience, the knowledge and the expertise to do that, we don't. And it is a pilot for *net-archaeology*, which is what we are doing now and what we are talking about. So, the big challenge, you know... and you recognize it...

E: And this is where we're standing. This is a typical drawing, like you know... like in the garage, in the café, we draw pictures like this, how we are going to do... And I work with *Paul Vogel*, and he is a very good, intelligent, and very smart guy and he is the technology guy. So, he has written some scripts. Those scripts are part of the name of what we call the "*Archie*". And we are making the over source. We are not finished yet, so we cannot promise anything. But we are aiming to finish and put it online.

C: It is really great plan.

E: What we really would like is... not that *Digital City* is virtual, and it's sunken and it is not to be seen, but it is *virtual Pompeii*. That we have reconstructed as much, and you can... maybe you will see ruins, but we hope we can show you something that we called *Fort Collection*, because you have height, density and time. Because we have the three interfaces, although different work system we will fall out of that... But we will have many points of time of the City. And that you can compare with *web-archive* from *Brewster Kahle*, because what he does is *snapshots*, so we are not sure how it will look, but we hope you can zoom in and just switch between the times. And then you have *Fort Collections*, because you add the time to...

C: It is a nice concept as well.

E: Yes. You're very happy. We just need a lot of hands to help us. So, what we do, we have five stages, we're still in the second and third. We started already two years ago. Three years ago with the planning, of course. And we have launched the first website, which we're transforming into the Museum corporation platform. We have a *Grave Diggers Party*; where we collect a lot of... and that is the *crowdsourcing*... we collect a lot of material, physical and virtual. I'll show you later. Now, we are *crowdsourcing* for stories and of course for pieces, either is a paper, handbook or screenshot you have digital.

C: May I ask you this... Is this *crowdsourcing* on the parties of that event, but also online? Because, I told you I went online, but I was not able to discover all the features in the website. So, I would like also to ask at this interview, how did you *crowdsource* for this?

E: We are now as well thinking in how we're going to activate it more. Because we do ask people: please, tell us your story, and we do that by social media. And I also do this one day a week. So, it is not that we are doing it all the time. So, we really pinpoint and focus on the... but we'll start doing this more, because now we have put it in the Museum, now we would like to put the *Fort Collection* on the chair. Now we have video-feeding, also very nice. And of course we want more. So, in that process we'll start collecting more, as well, of the stories. But we do ask people, we do talk and I do talk to people, I do give presentations here and there. So, we ask people to: send us in and give us your material.

GRAVE DIGGERS PARTY: A PARTY WITH A CAUSE

Crowdsourcen van digitaal cultureel erfgoed

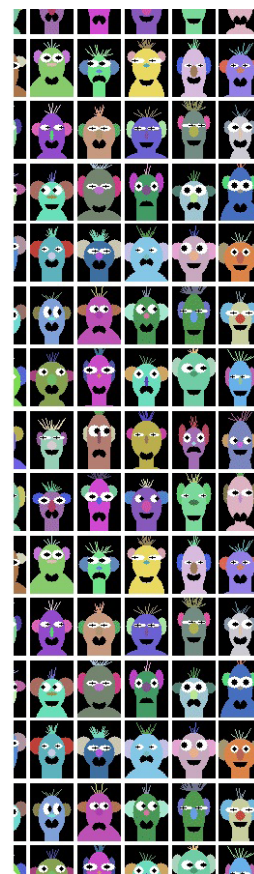
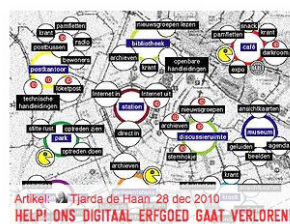
Ik ben fan [Tweet](#) [Vind ik leuk](#)

Vrijdag 13 mei 2011 vond de Grave Diggers Party plaats in de Waag Society te Amsterdam en daarmee de aftrap van de web-archeologisch werkzaamheden. Oud-medewerkers, oud-bewoners en web-archeologen van De Digitale Stad waren uitgenodigd om verloren herinneringen, zowel persoonlijk als digitaal, op te halen en langs te brengen. Het was een groot succes: er zijn vele floppies, schijven en servers opgegraven. Hiermee zijn de eerste stappen gezet voor het project re:DDS, waarmee we hopen DDS te kunnen reconstrueren.



@reneKorff
/at grappig: De opkomst en ondergang in de Digitale Stad: Uitgesproken VARA itgesproken.vara.nl/Nieuws-detail.... Dat echat is de bruine kroeg was leuk :-)

Artikel: Tjarda de Haan 2011
VARA: DE DIGITALE STAD ALS MUSEUMSTUK



And we have got a lot of material, which we are describing and putting as a *Functiekaart*, you know. Any idea how we call it in English? You know, the animation card, like in the Museum, you know. I can show you. So, every time we find something we describe it 27:27 [ruído]. So, we want let the bikes free, of course and share the knowledge. So, this was the *Grave Diggers Party*.

A lot of people we knew, like old system administrator from the *Digital City*, we knew we invited everyone who wanted to come, because there was also beer and reunion... Here you see if you brought something in, because that was the question: please, bring in your goods. Then we would write it down. Because we knew where did come from. So, we had work stations, you know, with the red and white tape, like police forensic scenario. And we had all the equipment, so people came with floppy or disc, and that moment we put at our discs, so they could bring their own disc back home. So, we had a lot of tools, and I also gave tours around what we've found. Like a museum guide tour. We also had workshops to test our scripting, and we still testing our scripting. And we got really nice feedback. So, we'll succeed in the exposition at the Museum.

E: So, we have at least... this is the collection that I showed you about it. And, well, we were talking about it, how we will continue the time machine, for the collection where you can switch. We hope, we are working hard for it. It can also be a playground for researchers, because actually the whole city is a source of historical documentation. How were people discussing? What did they think about Europe? What kind of documents were there? What was the first sent e-mail in the Digital City?

C: And I have another question also about this. Because I am very interested in this as a historical source and I do think it's really important. How will the Museum, as institution, make it available or not, the content that were at that moment private?

E: We are very clear about that. What was men's private, is private. This is a major issue. It is one the issues. Like, can we really dig up everything? How to deal with the complex material, like copies? What is the object? When it is an object? Because it's change every second, because somebody put in a cabinet, one of the articles. Where is the collection and what? And also the preference reasons. That is a very good start, if we go back then to the *freeze*. Because the *freeze* was taking in 96. Therefore, we are going to *freeze* everything. And then was a lot of discussion, within the people, who are going to make the *freeze*. And the decision was we are not putting the private sensible data on the tape. What we are going to put on the tape is software, the houses and manual of how we all did it, and manual how it all works, for the digital future archaeologist. So, on the *freeze*, and we're still unpacking the *freeze*, we haven't seen that yet. Officially, that shouldn't be any personal data on it, but of course, we in the *Grave Diggers Party* we got a lot of backups and in the backups there is information like that. So, we'll not just go to put on Get-up, we will really take that down. And during that process, we're thinking about it, and we are going to write it down, as a chapter in our handbook, because we will write the do-it yourself digital handbook for *web-archaeology*. Like other institutions can follow it, sees how we failed in many ways, but hopefully also how we succeeded. And privacy will be

one of the chapters, because it is a major and important thing.

C: I am very glad to hear that. And I also would like, if I understood it's like you are going to make a viable... just to make sure if I understood. Like the interfaces says what was public at that time. So, what the user make public and so on... how his preferences or interactions that were public...

E: We will definitely going to discuss a lot about this subject, because there is a lot of difference between houses. Some houses might be more private than other houses. So, we are open to discuss that. So, there is also... first to find out what Google find, because we haven't reconstructed it yet. Well, privacy is privacy, and we'll take care of that. We will also talk with our partners, how they do it... because the *Royal Library* is archiving live websites. What we are doing is digging up the *dead-web*. And they are digging up the *live-web*. And they have *opt-outs*. They decided to *opt-out*. So, if you're going to get an e-mail if your website is archived. And then you might be really happy, because then your website is one of the cultural web-heritage. They should be archived. But they will tell you: we are archiving it, but if you don't wanted, please, let us know. But we are archiving it. So, they use this *opt-out*. That is one option to the old...

C: Yes, but then how you will find the users?

E: Exactly. There are so many questions. I'll keep you posted. This will be one chapter. Just to go further...the playground for the researchers is really incredible and historical document. Well, we hope to give to the collections, to the *Heritage Institute*, and then we will share our knowledge, because we love that... and we'll write this manual. And we are also writing down, the 23 things of web-archaeology with our partners. Because in the digging up and in the way we were working to reconstruct the city, we have encountered with a lot of colleagues. So, *Royal Library* has helped us really with thinking about archiving and they are helping us, hopefully, with the emulation of the city, because you need to reconstruct it and they also use web archiving technology. So, we are probably using it as well. They are helping us in that and they will write about how they are doing the *web-archiving*, and in a short way, but with links to their cool projects and hopefully fed with video... But, just you know... for everybody that it is interested, but not preferred. Not definitely the professionals. Also for professionals, but to give a summary and then I will write a little short story about how they helped us in this project. So how did we collaborate?

E: Partners are the *Royal Library*; *Digital Erfgoed Nederland*, they also an English citation. *Beeld en Geluid*, is helping us. A lot of partners of our society, like *Marleen Stikker* is our director and she is helping us with the project, and a lot... And that is very nice. So, they'll also

take the [ruído; palavra não identificada} and will divide things in 23 things.

E: So, sharing a lot of knowledge, and what did we find? Well, forty one of root data, so we are packing that out. And the *freeze* and I will show you some pictures now... and we have found computers, complete service and... Well, these are the digital things, the physical things... The chair we will see later on... And all the digital things we have found in the software.

E: The *freeze*. We've found on the web archiving, a page about what happened to the *freeze*. That was a birthday story. What did they pack into the *freeze*? So, that is really like a manual for us, like to rise. We found three sorts of tapes. Those were the tapes. So, we were finding a tape reader and in the *Royal Library*. I've asked them: please, do you have a tape reader? And it took a long time. Then, they wrote me yes. But, we didn't know, because it was used as 37:08 [ruído]... Welcome, here. And you can have the tape reader. But it was old and they also couldn't work with, thus it really function? So, we've tried it. And we have the original tape, but we didn't have the back-up tape. So, you are sitting like: should we try it? It's digital heritage. If you mixed or fucked it up, and get caught, so we tried it and we got stuck. We were like sweating. But then we got it out again, and then we tried it with other people from the 37:42 [ruído; palavra não identificada] the *Royal Library*. And then we met *Henk Peek*, from the *Computer Museum*. Henk is the hero, because he got eleven gigabyte of raw data out of the tapes. And they look like these. It looks like the sterling code. And we've also found e-mails. So, we could really unpack that... and what we've found? The avatars. So, it was sitting there and we were so incredible happy to find it. Thus, the history of the *freeze* started in 96. Now, 2014, in the Museum, we are trying to make this reconstruction. So, we are very happy, and as well with the management.

E: This is the GEM, *public terminal*. And you will found these things also online and I will show it later. A lot videos we have found, we have to put it online. Marleen Stikker got it on the one year birthday. Manuals, we have already put them online. Now, we are going to more technical stuff.

C: That is such a useful work.

E: yes, it is. This is the technical stuff. This is how it is written in manuals. And this is like the reconstruction, the forensic research. We take scripting concepts of many different disciplines, like from data we covered the forensic, and discover programme, and all different disciplines. And we are adding restoration and pointing time, and many more. And these are the greater decisions. How we are going to do it? I can explain to you. But...

C: No, this part is also alright.

E: And these are the twenty three things, and which are working with us... it is not clear yet. But *Europeana* wanted to work with this 39:36 [ruído], so we will publish those things in

the coming time. This was, well that is it.

C: And can you tell me just a little about the end of *DDS*.

E: The finishing?

C: Yes. How it was socially, I don't know... because I don't have much information about that, besides some articles in English.

E: Well, it started in 94...And it was like a pioneering project about interpreting what you could do with the cyberspace. What you can do with it? How we are going be able to involve the audience, the public and not only the government or military or the universities? So, how to involve everybody? So, it started as a public domain.

So, it has started as *maatschappelijk*, a societal... a social pioneering project. At the time in 2001, when it was taking online the landscape of internet, it has completely changed the internet. How you call it? 40:57 [ruído] and the digital city had before a lot of written code, some much other codes were written later with all the knowledge they had before. So, how you it... by door... in house. There are many companies who could deliver, cheaper, faster these things at the *Digital City* gave. There is a story that they did not want to put advertisements around the public city. The *Digital City* started its own commercial company, as well. And this was how to balance this situation. But then the management could not find a way how to survive, and it were a lot losses, egos bruised. You could say there was not enough entrepreneurship. There was not many... One of the important things and there were so many more companies, who could... better and even cheaper... There was administrated buyout, and the managers decided certain ways and not everybody did agree with that. And in the end they sold one part and a lot of people did not liked, because the city was started with public money. Now it was gone to the management, which did the management buyout, so they were officially in that. But we did like to buy the city... then later on the other companies with were part of the *Digital City* were stopped. And then they took the site offline in 2001. So, you can still find a lot of things in a way. Archive, of course. But it was taken offline. So, it is not really there anymore.

C: No, I tried to go through the way back machine, but I really haven't gone further. You just stop by the main error or we cannot go...

E: Yes, a lot of broken images. 43:32 [ruído]

E: Because when we were trying to do to this *Fort Collection*, to bring back the experience. So, you can really, feel how it was like back then. And of course it is not really fun to have to wait, like five minutes before one page comes up. Because that was as well, the experience of the time. I have a monitor of 144k and it was like if you're in a children's bike on a free highway. So, you really went slowly. It took time that the pages have loaded.

C: But don't you think to emulate this time is also something interesting?

E: Absolutely. The time experience...you should be there... but if you are making people wait five minutes for... It might stop people, as well as to experience it. You might want to play with that. First, we have to dig it up, ok?

C: OK. It is such an interesting subject. I would like to see, to research... to know more about it.

E: Should we go on?

C: Yes. Thank you!



Tjarda de Haan. Museu de Amsterdam.
Foto após a entrevista na visita guiada a coleção de arqueologia digital.

Entrevista com Gustavo Viberti, em 26 de junho de 2014.

Camila: Gustavo, agradeço a você por participar deste projeto de doutorado e colaborar com seu relato sobre a história de inovação do Cadê?. Eu queria começar perguntando como foi essa experiência de, a partir da sua página pessoal, criar uma das primeiras ferramentas de busca brasileira?

Gustavo: Quando a gente fala daquela época tem que tentar voltar um pouquinho ao contexto do momento. O Cadê? começou em 95, quando a internet no Brasil era muito pequenininha. O acesso à internet para o público em geral havia sido autorizado poucos meses antes. Se não me engano em junho ou julho de 95 a Embratel deixou de ter o monopólio sobre o fornecimento de acesso à internet e abriu para provedores. A partir de então algumas pequenas empresas fornecedoras começaram a prover acesso à internet e pessoas fora do meio acadêmico começaram a entrar e eu fui uma delas.

Eu tinha acesso a um BBS que foi o primeiro que se tornou provedor aqui no Rio. BBS era o avô da internet. Assim, eu comecei a usar a internet logo depois que ela foi, digamos, liberada para o público em geral aqui no Brasil. Naquele momento eu estava procurando o quê fazer... Trabalhava em uma empresa de informática, estava chegando aos 30 anos e eu com outros amigos imaginávamos que era uma boa época para começar a fazer alguma coisa por nossa conta. A gente conversou muito e resolvi: “eu vou tirar férias e tentar entender melhor o que fazer”. Foi justamente quando eu comecei a usar mais a internet, e, para minha surpresa, vi aquela coisa toda começando aqui, fora do meio acadêmico, e lá fora algumas coisas começando também e achei que valia a pena testar fazer alguma coisa com internet. E aí foi natural fazer uma busca porque era um momento de início. A gente via que alguma coisa seria necessária para que as pessoas soubessem o que era aquilo – o que era a internet? O Yahoo!, que era um exemplo lá fora, havia começado como empresa poucos meses antes com um pessoal que estava no trailer desde a faculdade. Eles saíram dois ou três meses antes e montaram a empresa. Então, a gente via aquilo tudo acontecendo nos Estados Unidos, alguma coisa na Europa e aí eu comecei a imaginar que aqui também poderia acontecer... O Brasil adota muito rapidamente certas tecnologias, não é?

Camila: Isso é verdade.

Gustavo: Bom, então começou com uma página pessoal como um teste, mas já imaginando que aquilo poderia dar certo. E se realmente isso acontecesse a gente poderia montar uma empresa nesta área.

Camila: Qual é a sua área de formação? Pergunto porque essa familiaridade com a internet, naquele momento, não era comum...

Gustavo: Não, não era... Eu sou engenheiro de sistemas, me formei em Ciências da Computação e trabalhava numa empresa multinacional de informática. Embora eu não trabalhasse em uma área tão técnica, (era consultor), eu, por exemplo, já usava e-mail dentro da empresa, embora não fosse chamado de internet. Estava acostumado a usar e-mail e a acessar grupos de discussão, tipo o que viria a ser a Easynet, através da empresa. E também a BBS era meu mundo. Mais tarde, conversando, as pessoas falavam: “é, vocês tiveram sorte, foram os primeiros”, mas não acho que era bem sorte, é que a gente estava naquilo. Bom, sorte foi que vimos que aquilo se encaixou muito bem no que conhecíamos, era nossa área. Encaixou-se com o que a gente conseguia imaginar que aconteceria com uma ferramenta daquelas.

Camila: O Alternex, por exemplo, que foi um dos primeiros provedores do Rio e era ligado ao Ibase, também tinha uma ferramenta: o BookMarks, a partir de 98. Mas, ao que parece ninguém o utilizava muito...

Gustavo: É, realmente. Se por um lado nós vimos uma oportunidade para começar, e éramos da área, por outro lado, muita gente da área também poderia ter visto. Talvez tenha faltado acreditar que aquela coisa ia crescer. O Cadê?, de fato, não foi a primeira ferramenta de busca aqui... Havia, como você falou, o BookMarks que o pessoal do Ibase tinha no Alternex. Na Unicamp também eu me lembro que havia um pessoal que havia começado a juntar as páginas, criado alguma coisa. A RNP lançou também uma que se chamava Yahi?. O rapaz responsável fazia um bom trabalho, foi uma pena depois a RNP não querer continuar, porque eu acho que ele poderia ter ido mais longe se tivesse mais respaldo.

Mas nós estivemos entre os primeiros. Eu acho que nós fomos os primeiros que ficamos conhecidos, realmente. E aí, talvez a diferença tenha sido que nós já não éramos tão garotos. Eu falo ‘nós’ porque tive um sócio, o Fábio, que era meu amigo e é meu amigo até hoje. Nós estávamos na mesma fase da vida, chegando aos 30 anos, então o Cadê? começou como uma coisa pessoal que a gente imaginava que poderia se transformar em uma coisa legal. Naquele tempo havia muita colaboração, as pessoas estavam muito propensas a colaborar com tudo. Então, por exemplo, jornalista, se você mandasse uma *press release* legal, razoável, sobre internet, ele publicava. Hoje, para você conseguir que alguém publique alguma coisa é muito difícil. Então, na época, fiz uma lista de todos os jornalistas que eu lia no Brasil, e mandava para eles toda semana uma *press release*, alguma coisa interessante que tivesse acontecido através do Cadê?. E as pessoas publicavam.

Camila: E você e Fábio Oliveira que faziam tudo nesse começo do Cadê? Por exemplo, escolher as categorias, fazer o modelo dos diretórios, - que seguia um pouco o modelo do Yahoo!, mas com as características próprias do Brasil -, tudo isso era feito por você e o Fábio Oliveira, só?!

Gustavo: É, por algum tempo aconteceu basicamente assim: eu fiquei dedicado a isso nos meus dias de férias e consegui começar e lançar o Cadê? durante esses dias. E aí neste momento, quando ele foi lançado, isso era outubro de 95, eu lembro que eram só 300 páginas brasileiras. Era muito pouco. O que fiz, então, foi pegar o que havia, o que vi no Yahoo! e o que eu vi num outro site italiano, também, que eu achava legal, e adaptar. Peguei as categorias e adaptei porque eram só 300 páginas, então não tinha necessidade de muita categorização. Então, quando começou era uma coisa muito pequena, havia categorias principais, que no início era uma dúzia, ou talvez dez, depois passou para dezesseis. Mas aí as pessoas começaram a ver e todo dia me mandavam páginas novas, no início eram 10 páginas, 20 ou 30 por dia. Isso foi em outubro. Já em dezembro eu recebia por dia umas 150 páginas... E eu trabalhava! Então começou a ficar complicado para mim. O Cadê?, inicialmente, era uma coisa muito pessoal, era muito simples, mas então a coisa começou a crescer tão rapidamente que eu conversei com o Fábio e ele entrou também no projeto. Nós contratamos o irmão do Fábio, que era um pouco mais novo e tinha tempo disponível, e ele passou a dedicar algumas horas do dia dele a isso. Ele foi o primeiro editor, realmente, do Cadê?, ele pegava as páginas e as incluía.

Camila: Em relação a este momento inicial, a estas 300 páginas, você usava basicamente o Yahoo! e o Altavista para poder pesquisar essas páginas?

Gustavo: Não, usava tudo! Peguei tudo de tudo e, inclusive, de uma revista que havia na época, a Internet Brasil, se não me engano, que publicava em cada número páginas amarelas ao final da edição. Eram páginas para o pessoal entrar, sites que as pessoas mandavam. Pegava a revista e o que saía anotava, enfim, colocava tudo o que encontrava. Mas, em pouco tempo, as pessoas começaram a mandar os pedidos. Então, uma coisa que, inicialmente eu pegava ali fontes diversas para começar essas 300, já logo depois, poucas semanas depois, já não precisava mais porque as pessoas, quando faziam suas páginas, mandavam. Passou a ser uma coisa das pessoas fornecendo conteúdo, que é uma característica ligada à internet. Então as pessoas começaram a enviar através de e-mail, no início, as inclusões. Isso era no início, mas em dois ou três meses eu já não conseguia fazer... Foi quando entrou o Claudio (irmão do Fábio). Então a gente começou a ter um custo, porque tinha que pagar uma pessoa... Este início foi muito interessante, porque – isso foi há tanto tempo atrás, mas mesmo naquele momento – na internet, quando alguma coisa dá certo, dá certo muito rápido! Então, por exemplo, era uma página pessoal, que estava hospedada no meu provedor de acesso. Mas, um mês depois que ele começou, o pessoal do provedor me ligou e falou: “olha, está havendo um problema aqui porque a tua página pessoal sozinha consome mais do que todas as outras juntas. O servidor está caindo, não está aguentando...”. Aí me chamaram e eu pensei: “vão querer que eu saia do provedor”.

Mas, na verdade, eles gostaram da coisa e quiseram colaborar. Disseram: “vamos tirar você daqui e colocar sozinho num servidor só para o Cadê?”. Então, colocaram sem cobrar. Porque também, se eles cobrassem, nem teria como pagar. Naquele momento tudo isso era muito caro! Um mês depois eles me ligaram de novo: “olha, a gente gosta muito de você, mas o nosso canal fica comprometido porque você está consumindo já um volume importante... Pode ficar mais um tempinho, mas já vai pensando em como fazer, porque você vai ter que sair daqui a pouco”. Então, ao mesmo tempo em que era um problema, também era uma coisa legal! Porque você via que o volume estava crescendo muito, as páginas cresciam, as pessoas cresciam, eu acompanhava as visitas diárias que também cresciam muito. Então, a imprensa começou a dar atenção e, por causa disto, três meses depois que o Cadê? foi lançado, a gente teve um primeiro contato de gente interessada em investimento... E não se imaginava uma coisa assim no Brasil, naquele momento... A internet não era importante, nem lá fora era muito importante! Os IPOs ainda não tinham começado, então estava um pouco fora da nossa expectativa conseguir um investidor em tão pouco tempo. Imaginávamos que faríamos as coisas com o nosso dinheiro, só que os custos tornavam isso inviável, o custo de acesso à internet era muito caro... Felizmente, nós tivemos um pessoal interessado em entrar como sócio no negócio e a nossa primeira conversa com eles foi em janeiro. Então, foi tudo muito rápido. Quando uma coisa dá certo, ela é muito rápida!

You are viewing an archived web page, collected at the request of [Universidade Federal do Estado do Rio de Janeiro](#) using [Archive-It](#). This page was captured on 0:54:33 Dec 25, 1996, and is part of the [Brazilian Web Engines \(1997-2013\)](#) collection. The information on this web page may be out of date. See [All versions](#) of this archived page. [Metadata](#)
[Enable QA](#)

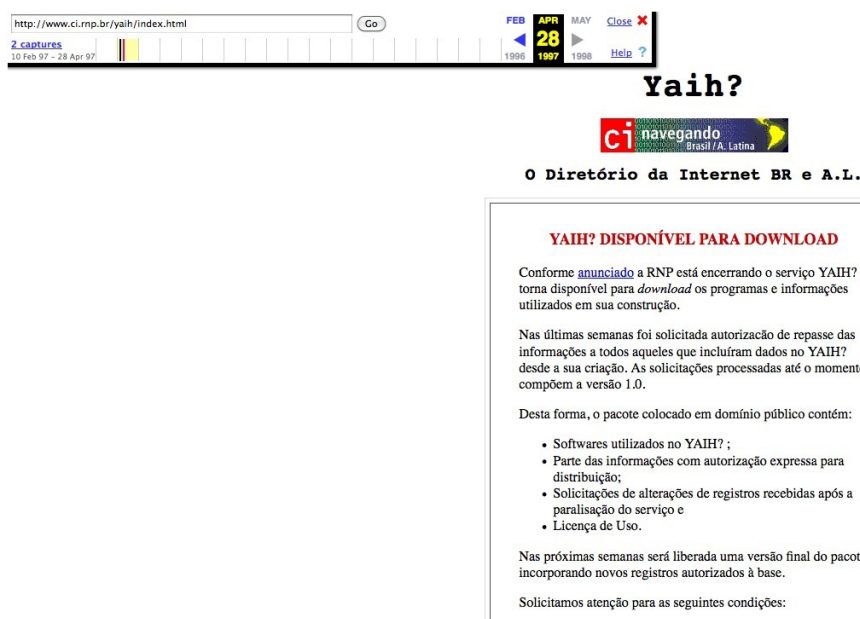
Camila: E nesse momento vocês tinham relações com outras pessoas que estavam neste mesmo setor no Brasil, pessoas que faziam buscas?

Gustavo: Na verdade não. A gente, no início pelo menos, se concentrou completamente no nosso problema e não tínhamos contato com ninguém da área de busca. Mesmo porque o único que eu via fazendo um trabalho legal era o pessoal da RNP, e era um rapaz que traba-

lhava sozinho!

Camila: E nesta época o Cadê? teve um crescimento exponencial visível! Pergunto se a área de conteúdo, logo em seguida, passou a ter pessoas responsáveis por essas categorias ou se manteve, basicamente, a estrutura inicial?

Gustavo: Quando a gente começou a crescer e a internet no Brasil também, nosso critério para inclusão eram páginas em português, não necessariamente do Brasil. Então, se tinha alguma coisa interessante em português de Portugal, a gente incluía também. Mas o volume aqui, de qualquer maneira, era muito maior do que de Portugal e cresceu muito rápido. Então, o Cláudio, que nós tínhamos contratado para trabalhar algumas horas do seu tempo, logo depois foi contratado *full time*. A gente abriu um escritório, então passamos a ter um local externo de trabalho, em vez de ser a nossa casa. Isso foi junho de 96. O Claudio era o cara que incluía e, quando tinha que incluir uma categoria nova importante, ele me consultava e a gente via se era necessário ou não. O critério, basicamente, era: se nós tivéssemos várias páginas, mais do que três, que ficassem melhor em uma nova categoria inicial, nós abríamos. Se não houvesse pelo menos três páginas, elas ficavam numa categoria inicial mais generalista. Uma nova que a gente incluiu no início, que o Yahoo! não tinha, mas para cá era muito relevante, era 'compras on-line'. Sites que vendiam pela internet, que no início havia pouquíssimos. Nas categorias de baixo, a gente criava à medida que iam surgindo pelo menos três páginas que tratassem daquele assunto. Então, no início o Cláudio me consultava, mas depois ele passou a ter experiência e ele mesmo fazia isso sozinho. Depois contratamos outro editor, e eu fiquei no Cadê? até 2000. Nesta época já eram quatro editores, um deles com especialização em biblioteconomia, para dar uma ajuda.



Yaih?

Ci navegando
Brasil / A. Latina

O Diretório da Internet BR e A.L.

YAIH? DISPONÍVEL PARA DOWNLOAD

Conforme [anunciado](#) a RNP está encerrando o serviço YAIH? e torna disponível para *download* os programas e informações utilizados em sua construção.

Nas últimas semanas foi solicitada autorização de repasse das informações a todos aqueles que incluíram dados no YAIH? desde a sua criação. As solicitações processadas até o momento compõem a versão 1.0.

Desta forma, o pacote colocado em domínio público contém:

- Softwares utilizados no YAIH? ;
- Parte das informações com autorização expressa para distribuição;
- Solicitações de alterações de registros recebidas após a paralisação do serviço e
- Licença de Uso.

Nas próximas semanas será liberada uma versão final do pacote incorporando novos registros autorizados à base.

Solicitamos atenção para as seguintes condições:

Camila: No acervo do Internet Archive, que é o único arquivo on-line que coletou dessa época da web, ele começa as páginas do Cadê? em 97. Nesta época o Cadê? tinha 16 páginas enquanto que o Yahoo! tinha 14. Antes desta época não está on-line, eu não tive acesso...

Gustavo: Sobre essa preservação da memória de informação digital, eu lancei uns livros, junto com o Cadê?. Naquela época, nem todo mundo acessava a internet, mas estava querendo acessar e uma editora me ligou e falou: “olha, por que você não lança um livro de páginas amarelas?”. Primeiro achei que era uma maluquice, mas depois pensei que poderia ser interessante até para que as pessoas conhecessem melhor o Cadê?. Então, tem um livro de 96 que, basicamente, é o Cadê? da época, chamado Páginas Amarelas . Então, tem essa primeira edição que mostra bem as páginas iniciais da internet no Brasil.

Camila: Quais eram os filtros que você lembra que usava na época para as páginas que não eram colocadas?

Gustavo: Isso foi acontecendo ao longo do tempo, a gente foi criando regras de edição. Por exemplo, uma que de cara a gente criou foi ‘pedofilia’. Quando uma pessoa submetia uma página, os editores entravam na página para dar uma olhada e fazer uma resenha dela. Era um trabalho que demandava algum tempo, mas que evitava que entrassem no Cadê? propagandas e coisas assim. Então, toda resenha era feita pelos editores. Quando eles entravam na página, viam se realmente se tratava daquilo proposto. Se houvesse pornografia infantil, a gente denunciava a página. Tínhamos um canal com a Polícia Federal, um grupo da PF que trata pedofilia, e a gente enviava para eles. E não incluía, claro. Outra era ‘terrorismo’. Uma vez apareceu uma página que ensinava a fazer bombas. Discutíamos sobre liberdade na internet, conteúdo livre, ideias livres, mas quando chegou uma página deste tipo, que ensinava como fazer uma bomba, não dava para incluir. Então, a gente procurava ter critérios mínimos para, por um lado, não coibir a liberdade daquilo que fosse criado, mas, por outro lado, coisas como pedofilia, ensinar a fazer bomba, não entravam.

Camila: E em relação ao grau, ao ranking das páginas? Era feito imediatamente quando os editores colocavam essa pequena resenha ou à medida que as buscas iam se processando se fazia alguma modificação?

Gustavo: O Cadê? funcionava como um catálogo, então, na verdade, a gente não mudava. No momento da busca a gente não alterava uma classificação em função da popularidade da página. Quando você fazia uma busca ele procurava aquilo que se buscou na resenha, no título ou nas categorias da página e te mostrava, sem mostrar em primeiro lugar uma página ou outra

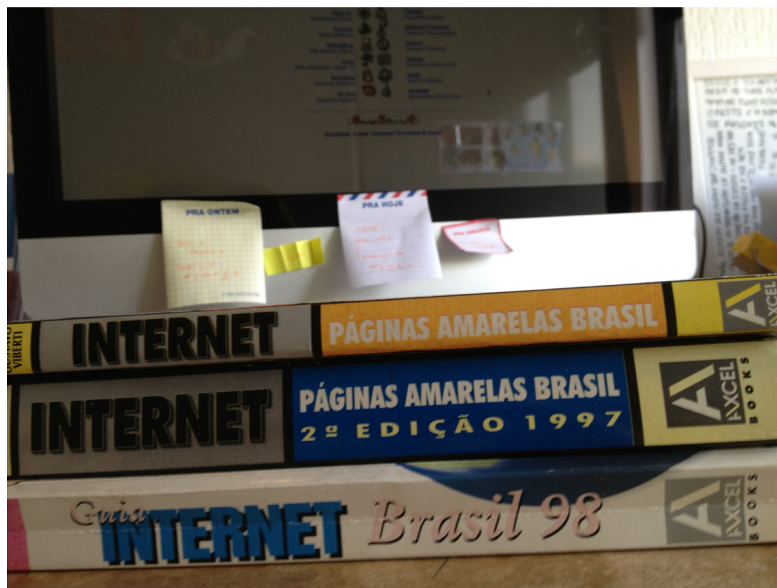
porque era mais popular ou não. Lembra, isso aí aconteceu antes do Yahoo!, antes do Google...

Camila: Antes do Google, sim.

Gustavo: Então, é um mundo muito diferente. O Altavista existia e fazia alguma coisa assim. Eu lembro até quando o nosso gerente de tecnologia, lá em 98, por aí, me mostrou o Google, ainda como página do pessoal da faculdade de Stanford, e o algoritmo publicado, que era o trabalho de doutorado deles. Lembro que a gente leu aquilo e achou ótimo, uma coisa legal, mas a gente estava em um momento (98 para 99) que a empresa estava sendo vendida... Esse nosso gerente de tecnologia, que é um cara muito competente e que hoje é o responsável pelo Bing, da Microsoft no Brasil, queria seguir por este caminho (o da Google), mas naquele momento a gente não podia se enveredar por projetos novos. Então a classificação era feita mesmo somente por categoria e, dentro da categoria, por ordem alfabética.

Camila: Vocês salvaram um pouco desses dados de servidores ou você tem apenas os livros mesmo?

Gustavo: A gente se preocupou, naquele momento, com cópias, porque o Cadê? era muito copiado. O nosso pessoal fazia um trabalho editorial, entrava nas páginas e fazia uma resenha sobre o conteúdo, tinha todo este trabalho... então a gente ficava chateado quando via aquela resenha igualzinha em outro site de busca. Então, em algum momento isso nos incomodou e a gente procurou uma forma de proteção. Não havia nada na legislação brasileira. A única solução que o advogado editorial deu, na época, foi registrar como livro na Biblioteca Nacional. A gente fez uma listagem de tudo o que tinha na Cadê?, em 96 ou 97 – acho, e imprimiu tudo. Era uma listagem muito grande, muito grande. Imprimimos e registramos na Biblioteca Nacional como livro, como catálogo. Então, existe essa listagem do Cadê? em algum lugar da Biblioteca Nacional, onde eles guardam essas coisas.



Catálogos impressos, gentilmente emprestados por Gustavo Viberti .
(Na tela, o cadê no Internet Archive).

Camila: Interessante saber isso porque mostra, de um lado, a falta de regulamentação naquele momento, e, de outro, a solução de transpor para o meio impresso...

Gustavo: Exatamente! Por um lado o meio impresso garante a memória, não é? Então, este documento pode mostrar como é que era a internet brasileira em 96 ou 97, o que havia ali, as páginas incluídas... Naquela época ainda dava para imprimir, mas pouco tempo depois já não daria mais porque ele cresceu tão rápido! Seria um problema tentar imprimir, dariam caixas e caixas de papel.

Camila: E como foi essa decisão de vender empresa? Vocês ficaram? Ela foi vendida para um grupo que permaneceu? No final ela foi vendida para o Yahoo! não é isso?

Gustavo: Ela foi vendida primeiramente para a Starmedia, que foi uma empresa que teve um sucesso muito grande. Era uma empresa americana, de Nova Iorque, que tinha como objetivo ser a porta de entrada na internet para os latino-americanos. Uma coisa, assim, que para um americano fazia algum sentido, mas para nós não fazia muito, mas enfim... Era tão efervescente esse mercado de capital de risco nos Estados Unidos, que esse pessoal da Starmedia conseguiu muito dinheiro com essa ideia. Eles levantaram um bom dinheiro quando fizeram o IPO na bolsa e foram comprando empresas. No Chile, México, Espanha e também compraram empresas no Brasil. Aqui eles compraram um mecanismo de busca que se chamava Zeek, que era um mecanismo de busca que existia na época, não era muito grande, mas era um dos que naquele momento ainda existiam. Naquele momento decidimos vender a empresa porque havia muitas ameaças externas ao nosso capital. A empresa tinha crescido, mas não era tão grande assim para fazer frente à gente muito grande

que estava chegando ao Brasil naquele momento. Primeiro começamos a procurar um investidor, alguém que entrasse com algum dinheiro para fazer a empresa crescer mais rápido, mas pessoal não queria isso, eles queriam realmente comprar tudo. Então eles compraram a empresa em 99 e nós permanecemos ainda à frente dela até 2000, durante um ano depois da compra, por uma condição contratual que os compradores colocaram. Durante este tempo, o Cadê? continuou operando como sempre trabalhou, só que com algum dinheiro a mais. Não muito a mais, porque, na verdade, o dinheiro que chegava deles era mais ou menos para propaganda em televisão. Algo diferente do que a gente faria. Em 99 a empresa já tinha um faturamento importante e conseguia crescer, não tão rápido quanto a gente gostaria, mas de qualquer maneira, começou a crescer bastante. Quando eu saí da empresa éramos 50 pessoas. O crescimento foi rápido porque a propaganda na internet havia começado a ganhar importância e o nosso objetivo no Cadê? era vender publicidade.

Camila: E essa publicidade eram mais banners ou era também a própria resenha de sites?

Gustavo: Não, não, a gente separava uma coisa da outra. No lado editorial, a gente não cobrava por inclusão da resenha, não vendia isso, é como um jornal, não é?

Camila: É a internet antes do Google...

Gustavo: É, exatamente. A gente não fazia aquela coisa de colocar na frente, não mexia em nada disso. A gente queria preservar uma credibilidade editorial, que era o conteúdo. Naquele momento também, como não havia tanta propaganda assim, não tinha tanta coisa, então, os banners tinham um resultado razoável, um resultado melhor do que depois passaram a ter. As pessoas clicavam lá. Para você ter uma ideia, nosso primeiro cliente importante, em 96, foi o Ministério da Educação, com uma campanha feita pela McCann & Ericson, do Enem. Na verdade, na época chamava Provão, que havia começado em 95, se não me engano. Aquele, de 96, era o segundo Provão e eles queriam fazer propaganda nos lugares em que eles imaginavam que os jovens estariam... E aí foi a primeira propaganda importante, que veio de uma agência grande e que tinha uma verba legal. E para nós foi importante também entender qual era o resultado daquilo, entender como podíamos influenciar, ou seja, qual era o alcance que o Cadê? tinha. E foi muito legal porque o objetivo da propaganda era o seguinte: mostrar um banner que dizia para as pessoas que no dia tal, do mês tal, na hora tal, haveria um chat com o Ministro da Educação, onde as pessoas poderiam entrar e tirar dúvidas sobre como é que era o Provão. Então, era uma conversa com o Ministro através da internet, era uma coisa legal.

Camila: Realmente.

Gustavo: No dia do evento, da conversa com o ministro, nós ficamos esperando lá no escritório para ter um retorno do pessoal da agência sobre como é que estava. E aí o pessoal naquele dia não deu retorno. No dia seguinte, em um jornal aqui do Rio, apareceu uma foto do

ministro com várias pessoas em volta. Uma mesa com vários computadores e pessoas operando o chat, e eles diziam que o chat caía toda hora porque tinha gente demais entrando... Eles não esperavam tanta gente entrando naquilo e o Cadê? tinha sido o único veículo em que eles tinham feito propaganda. Então, foi uma das indicações legais, assim no início da coisa. A outra foi a Amazon que fez uma campanha, também no início de 96. Foi a primeira vez que a Amazon fez uma campanha fora dos Estados Unidos. Ela escolheu sete mercados, sete maiores, e o Brasil era um dos sete. O Brasil era um dos sete maiores mercados de venda de livros em inglês. E para nós foi muito legal. Além disso, a gente recebeu de pagamento um cheque assinado por Jeff Bezos!

Camila: Isso é histórico hein!

Gustavo: É... você me fez lembrar coisas que aconteceram há muito tempo... Depois coisas grandes aconteceram à empresa, mas aquilo que acontece no início, que te salva... chegou o cheque do Jeff Bezos de cinco mil dólares, para nós era muito dinheiro, cinco mil dólares! Eram coisas boas que iam acontecendo e que nos indicavam que o caminho era bom. Indicavam que a internet estava crescendo e que nós crescíamos junto, foi uma boa experiência.

Camila: Achei interessante você falar desta campanha do Ministério da Educação, até pioneiro...

Gustavo: É, às vezes, você imagina que o Governo demora, mas a agência estava muito atendida. Foi uma coisa interessante.

Camila: Você participou de algum órgão ou de alguma associação que visasse pensar sobre a regulamentação de algumas coisas da internet naquele momento, como RNP ou com o início do Comitê Gestor da Internet?

Gustavo: Com a RNP nós tivemos um relacionamento muito próximo porque, conforme eu estava contando aqui, a gente começou primeiro com um provedor, que era meu pessoal e depois de algum tempo tivemos que sair de lá porque aquele provedor não ia aguentar mais. Naquele momento ter um site de alto tráfego era muito difícil porque os provedores eram pequenos. Eles não tinham a banda muito grande e mesmo que tivessem seria muito cara. Então a gente começou a procurar um lugar para hospedar o Cadê? e acabou indo para o Laboratório Nacional de Computação Científica (LNCC), aqui no Rio, ligado à RNP. Nós conhecemos pessoas lá, apresentamos o projeto. Naquela época todo mundo achava isso legal, achava que deveria ter alguma coisa no Brasil assim e queria dar uma força. A gente acabou indo para dentro do LNCC, durante algum tempo, enquanto o Cadê? ainda não era uma coisa muito comercial, e ficamos lá alguns meses. Então isso aí criou um vínculo, também com o pessoal do LNCC, especialmente, que era o nó da RNP aqui no Rio. O Cadê? acabou indo para

a Embratel logo depois. Então a gente conversava com as pessoas que operavam a parte de infraestrutura porque a gente precisava disso. Conversava com as agências porque elas eram as intermediadoras da venda de publicidade. Agora, como órgão a gente sentiu necessidade de criar alguma coisa. Nós e outros veículos também como, por exemplo, o Universo Online, o Terra e outros que passaram a vender publicidade. A gente sentiu necessidade de criar um órgão que representasse este novo mercado. E aí se criou a Associação de Mídia Interativa, AMI, acho que foi em 98. Era uma associação cujo papel seria mostrar para o mercado essa área de internet como veículo publicitário, mas também criar para essa área regras, normas, métricas... Então na associação havia os veículos, as agências de publicidade e os anunciantes e esses atores tinham um papel na diretoria. Eu era um dos diretores e a minha responsabilidade, por exemplo, era criar as métricas para medir resultado de campanhas na internet. Como seria mostrado, como seria vendido, como seria cobrado...

Camila: Já é mais um dado importante para esse cenário: os atores que estavam ali, naquele momento, construindo essas primeiras normas, esses primeiros documentos...

Gustavo: Você falou com o pessoal do Ibase, da Alternex? Eles foram os primeiros, realmente, a trazer a internet para cá. O Carlos Afonso...

Camila: O Carlos Afonso tem muita coisa publicada, então dele eu consegui vários textos, artigos, o que foi bom. A pesquisa centrou um pouco nas ferramentas de busca, mas o Carlos Afonso é fundamental nesta história. Agora, realmente, sobre as ferramentas de busca do Brasil, logo depois, com a ascensão do Google, eu não encontrei nenhum estudo sobre isso. A não ser da área de computação, que tem muitos artigos técnicos sobre ranking...

Gustavo: O que eu posso te falar do mercado daquele momento: logo depois que nós começamos estava o pessoal da RNP com o Yahi?. Ele (o responsável pelo Yahi?) trabalhou muito bem nisso durante um bom tempo. E se eu fosse apontar alguém que fez concorrência ao Cadê? foi ele. No início éramos nós e ele, mas o projeto dele não andou. A RNP se desinteressou pelo projeto e sei que ele saiu dali e foi fazer uma ONG depois. Então depois de algum tempo nós não tínhamos concorrência. Era mais ou menos como se vê hoje no Google aqui... Talvez 90% das buscas no Brasil são feitas pelo Google hoje. Se você considerar os mecanismos de busca brasileiros naquele momento, não considerando os internacionais, nós tínhamos alguma coisa parecida: o Cadê? era 90% do que o pessoal buscava aqui, fora o Yahoo! para essas coisas internacionais. Eu cheguei a conhecer o pessoal do Zeek, que foi o segundo. Era bem diferente, uma coisa bem menor. A gente acabou realmente crescendo e conseguindo tomar o mercado, não digo nem porque era uma qualidade maior, não sei, acho que tivemos mais gente trabalhando nisso.

Camila: Eu acho que também estar fora de um estatuto governamental fez de vocês uma página mais próxima da cultura da internet. Porque tanto o Yahi?, quanto o Radix, em Pernambuco, em 99, e depois o TodoBr, todos eles eram muito ligados à universidade e tinham um traço muito acadêmico. Eu acho que o Cadê? não tem este traço tão forte acadêmico, o que aproximou mais as pessoas, talvez.

Gustavo: O que nós tivemos foi aquela coisa da empresa criada com pouco dinheiro, em que nós tínhamos que sobreviver e nos virar. Então nossa atenção toda estava voltada para aquilo. Eu tive um filho na época, que nasceu em 96, e tive muito pouco contato nos três primeiros anos de vida dele por conta do trabalho. Foi muito intenso.

Camila: Só para a gente concluir cronologicamente: você ficou até 2000 trabalhando no Cadê? e eu sei que depois o Cadê? foi vendido para o Yahoo!. Foi isso?

Gustavo: Isso. A empresa que comprou o Cadê? foi a Starmedia que é uma empresa que teve um sucesso muito grande no seu lançamento na bolsa, na Nasdaq. Mas o produto que eles tinham não era tão bom quanto eles imaginavam. Então, eles acabaram comprando outras empresas, e o Cadê? foi uma delas. Mas eles não conseguiram juntar essa coisa toda em um guarda-chuva como imaginavam, não andou. Então a empresa teve dificuldades e pouco tempo depois que nos comprou, não sei se foi um ano ou dois depois, basicamente o dinheiro dela acabou e ela se desfez. Antes de se desfazer ela vendeu algumas que tinha comprado. O Cadê? foi vendido, então, para o Yahoo!.

Camila: E depois disso você trabalhou ainda com alguma coisa relacionada com essas ferramentas de busca ou foi para outra área?

Gustavo: Não, depois que eu saí do Cadê? fui trabalhar durante algum tempo na área de capital de risco na internet. Eu era associado a um banco e a gente tinha ideia de criar no Brasil o que não havia, o que a gente não teve. Teve um pouquinho, mas o Brasil não se beneficiou daquela febre, que os Estados Unidos se beneficiaram, antes da queda da Nasdaq, de capital de risco abundante. Aqui não houve isso, então parecia um bom momento para começar, 2000. E eu entrei nisso, fizemos um grupo com banco, criamos uma incubadora, uma coisa por aí. Mas logo depois a Nasdaq caiu, a bolsa lá despencou e o mundo inteiro ficou durante alguns anos sem essa abundância de capital de risco.

Camila: E sobre essa história com preservação de memórias na internet, qual é sua visão sobre isso?

Gustavo: Eu estava pensando nisso há algum tempo atrás. Eu tenho filhos. Além de dois pequenos, tenho um filho, aquele que nasceu quando eu estava começando o Cadê?, que vai fazer 18 anos. E o que a gente passou naquela época é uma história legal pra contar para este

meu filho de 18 anos que vai entrar na faculdade de engenharia. Eu acho que ele pode passar por alguma coisa parecida, então, é uma coisa legal de contar. Acho que interessaria para algumas pessoas conhecerem um pouco mais como começou a coisa aqui. A história do Ibase é uma história legal; do Alternex, a história da gente e de outros. Eu acho que essas histórias estão preservadas de formas desconexas. Aqui com um artigo, ali com outro. Já se escreveram alguns livros de empreendedores e tal. Já fui entrevistado para algumas coisas assim, mas eu acho que falta alguma coisa um pouco mais organizada. O que o Internet Archive faz é muito legal: preservar aqueles momentos, a fotografia de uma época.

Camila: E existem em outros países estas iniciativas, como o arquivo da web portuguesa, por exemplo, que é bem acessível e muito legal, mas realmente o Brasil ainda não tem... É uma questão em aberto. Eu, pessoalmente, acho que é importante pelo menos ter um registro para se pensar. Histórias que merecem ser contadas e pensadas... A gente esquece um pouco, penso que os jovens podem achar que é só Google a internet...

Gustavo: É exatamente. Há pouco tempo se achava que era só Yahoo também, então...

Camila: É verdade e é interessante ouvir isso. Houve um tempo que também era só Altavista para algumas pessoas...

Gustavo: Exatamente, essas coisas mudam. Eu acho que o pessoal do Google foi incrivelmente competente. Eles tiveram sucesso não apenas pela busca, mas por várias outras formas, como o trabalho com a publicidade. Isso foi muito legal, mas nada garante que daqui a dez anos as pessoas vão continuar também se lembrando da mesma forma do Google. Essa é uma indústria em que as coisas mudam.

Camila: Tem uma dinâmica interessante porque por um lado tem uma nostalgia, existem muitos sites que contam a história da internet, mas por outro lado existe pouca documentação real sobre o assunto. É mais fácil você encontrar um manuscrito medieval completo do que uma página da década de noventa.

Gustavo: É verdade. Com software também é assim. Quando eu comecei a usar o editor de texto, há alguns anos atrás, não era o Word que se usava, era outro, era o WordStar. Este era o mais popular, e ninguém lembra mais dele... O Excel também não foi a primeira planilha, havia outras. Há pouco tempo atrás eu vi um documentário que mostrava a história das planilhas e achei muito legal. É legal, mas ninguém se lembra disso e todo mundo usa o Excel...

Camila: Exatamente, faz parte do cotidiano...

Gustavo: É uma história muito recente. Eu acho que é uma área que, realmente, não preserva a história porque, talvez, as pessoas não se interessem muito em entender de onde veio aquilo. Talvez queiram saber o que vem depois... mas, enfim, eu gosto muito de História, então,

acho que vale a pena. Uma vez eu fiz uma entrevista com o Carlos Afonso para uma revista digital que lançamos no Cadê?, que se chamava Aqui. A pauta era a história da internet no Brasil. Quando os jornalistas que contratamos para a revista disseram que iam entrevistar o Carlos Afonso eu disse “quero ir junto!”. Foi muito legal ele contando como foi o início em 92. Acho que vale a pena preservar histórias para as pessoas saberem como foi que tudo aconteceu...

Camila: Esse é um dos objetivos do meu trabalho: que volte ao público também. Que volte a circular essa informação de alguma forma. Gustavo eu te agradeço muitíssimo pelo seu tempo, pela sua disponibilidade e pelas fontes e informações que você está provendo para a pesquisa.

Gustavo: Eu é que agradeço a chance de relembrar um pouco.

Brazilian Historical Web Sphere (1997-2005)
Project proposal to the Internet Archive

Camila G. Dantas

Federal State University of Rio de Janeiro (UNIRIO)/Radboud University

June, 2013

Summary

How the Brazilian historical web sphere looked like between 1997 and 2005? What was being classified as historical web sites? There is lack of web-archiving initiatives by the Brazilian institutions such as National Library and National Archive (Dantas, Dodebei, 2010) what leads us to propose this partnership. It is our goal to build a special digital collection from the Brazilian domain that will remain open and accessible to the public. We propose to start with a case study of the Brazilian Historical Web Sphere, which will be mapped according to the native web classifications of the local search engines of the period.

Methodology

Following the works of Richard Rogers (2010) we are going to crawl our start points from the directory of the Brazilian web search engines. The idea is to map what was, from 1997 to 2005, being classified under the category of History. In doing this cartography it is also our goal to call attention for the history of local web search engines that were sold subsequently to Yahoo! And Google.

Proposed steps:

- 1.* To crawl the start points from the following web sites under the categories: History, Human Sciences, Library (História, Ciências Humanas e Bibliotecas)

www.cade.com.br (1997-2005)

<http://www.cade.com.br/cultura.htm>

www.radix.com.br (1999 -2005)

<http://www.radix.com.br/busca/Categoria?pagina=1&catg=cienciassociai1.Co><http://web.archive.org/web/20011215145105/http://www.radix.com.br/busca/Categoria?pagina=1&catg=hi>
[storia8.Co](http://www.storia8.Co)

www.uol.com.br (1996-2005)

www.todobr.com.br

2. To build an in-link network analysis in order to find out the center or relations of this web sphere.
3. To make an annotated web collection of the 100 web sites of the Brazilian historical web sphere, with public access.

Prospective Project

This case study could be a first step to build a bigger ambitious collection of the Brazilian domain what certainly would give much more context to this research. It is our hope to make a partnership with the Brazilian National Library in the near future, and together with Internet Archive, to reconstruct the ghost Brazilian web.

Works Cited

ROGERS, Richard. (2010) Internet Research: The Question of Method—A Keynote Address from the YouTube and the 2008 Election Cycle in the United States Conference, *Journal of Information Technology & Politics*, 7:2-3,241-260, DOI: 10.1080/19331681003753438

DANTAS, Camila G. DODEBEI, Vera. Research notes on the the emerging concepts of born-digital heritage in Brazil. *Proceedings CATAAC*, Canada, 2010.

URLS BRAZILIAN WEB SPHERE

CADE.COM.BR

<https://web.archive.org/web/19961225005433/http://cade.com.br/>

<https://web.archive.org/web/19970612222904/http://www.cade.com.br/>

<https://web.archive.org/web/19980710044720/http://cade.com.br/>

<https://web.archive.org/web/19970612223131/http://www.cade.com.br/referen.htm>

<https://web.archive.org/web/19990218123610/http://www.cade.com.br/>

<https://web.archive.org/web/20000229191126/http://cade.com.br/>

<https://web.archive.org/web/20020802103435/http://br.cade.yahoo.com/>

<https://web.archive.org/web/20041205152633/http://br.yahoo.com/cade/dir04/?http://br.cade.dir.yahoo.com/Educacao/>

https://web.archive.org/web/20040503201543/http://br.yahoo.com/cade/dir03/?http://br.cade.dir.yahoo.com/Artes_e_Cultura/

https://web.archive.org/web/20041205153316/http://br.yahoo.com/cade/dir13/?http://br.cade.dir.yahoo.com/Fontes_de_Referencia/

<https://web.archive.org/web/20030728173112/http://br.cade.yahoo.com/>

<https://web.archive.org/web/20030728173112/http://br.cade.yahoo.com/>

<https://web.archive.org/web/20040915042032/http://br.cade.yahoo.com/>

<https://web.archive.org/web/20050401090242/http://br.cade.yahoo.com/>

<https://web.archive.org/web/20060315132009/http://br.cade.yahoo.com/>

<https://web.archive.org/web/20070501111045/http://cade.search.yahoo.com/web>

<https://web.archive.org/web/20070501111045/http://cade.search.yahoo.com/web>

WITH IN-LINKS

<https://web.archive.org/web/19970612223131/http://www.cade.com.br/referen.htm>

<https://web.archive.org/web/19970612223028/http://www.cade.com.br/cultura.htm>

<https://web.archive.org/web/19970612223028/http://www.cade.com.br/cultura.htm>

<https://web.archive.org/web/19970612223028/http://www.cade.com.br/cultura.htm>

<https://web.archive.org/web/19971210075939/http://cade.com.br/cultura.htm>

<https://web.archive.org/web/20000815071301/http://www.cade.com.br/referen.htm>

<https://web.archive.org/web/20000815070708/http://www.cade.com.br/educacao.htm>

<https://web.archive.org/web/20000815070708/http://www.cade.com.br/educacao.htm>

<https://web.archive.org/web/20000816010518/http://www.cade.com.br/culhist.htm>

<https://web.archive.org/web/20000815071301/http://www.cade.com.br/referen.htm>

<https://web.archive.org/web/20000815070708/http://www.cade.com.br/educacao.htm>

<https://web.archive.org/web/20000816010518/http://www.cade.com.br/culhist.htm>

<https://web.archive.org/web/20000815070622/http://www.cade.com.br/cultura.htm>

<https://web.archive.org/web/20000816010518/http://www.cade.com.br/culhist.htm>

RADIX

<https://web.archive.org/web/19991128101010/http://www.radix.com.br/>

<https://web.archive.org/web/20001202015100/http://www.radix.com.br/busca/frequencia>

<https://web.archive.org/web/20000229231509/http://www.radix.com.br/busca/hits>

<https://web.archive.org/web/20011124025915/http://www.radix.com.br/busca/Indexht3>

<https://web.archive.org/web/20020405125642/http://radix.ibest.com.br/busca/Index>

<https://web.archive.org/web/20020407060237/http://radix.ibest.com.br/busca/Categoria?pagina=1&catg=educacaoerefel.Co&categoria=05>

<https://web.archive.org/web/20030421163429/http://radix.ibest.com.br/busca/Index/>

<https://web.archive.org/web/20030604023122/http://radix.ibest.com.br/busca/Categoria?pagina=1&catg=cienciassociai1.Co&categoria=09>

<https://web.archive.org/web/20041001011509/http://radix.ibest.com.br/busca/Index/>

<https://web.archive.org/web/20040722100036/http://radix.ibest.com.br/busca/Categoria?pagina=1&catg=cienciassociai1.Co&categoria=09>

WITH IN-LINKS

<https://web.archive.org/web/20011205013427/http://www.radix.com.br/busca/Categoria?pagina=1&catg=cienciassociai1.Co>

<https://web.archive.org/web/20011215145105/http://www.radix.com.br/busca/Categoria?pagina=1&catg=historia8.Co>

<https://web.archive.org/web/20011121004458/http://www.radix.com.br/busca/Categoria?pagina=2&catg=historia8.Co>

<https://web.archive.org/web/20020623043108/http://www.radix.com.br/busca/Categoria?pagina=3&catg=historia8.Co>

<https://web.archive.org/web/20020628203446/http://www.radix.com.br/busca/Categoria?pagina=4&catg=historia8.Co>

<https://web.archive.org/web/20011215131821/http://www.radix.com.br/busca/Categoria?pagina=1&catg=globalizacaol.Co>

<https://web.archive.org/web/20011215142448/http://www.radix.com.br/busca/Categoria?pagina=1&catg=cienciadainfo1.Co>

<https://web.archive.org/web/20020121050603/http://radix.ibest.com.br/busca/Categoria?pagina=1&catg=arteecultura4.Co&categoria=02>

TODO.BR

<https://web.archive.org/web/19991128213556/http://www.todobr.com.br/>

<https://web.archive.org/web/20001017114556/http://todobr.com.br/>

<https://web.archive.org/web/20000830142643/http://www.todobr.com.br/dados.php3>

<https://web.archive.org/web/20000830142659/http://www.todobr.com.br/tecnologia.php3>

<https://web.archive.org/web/20010330040955/http://todobr.com.br/>

<https://web.archive.org/web/20020401204157/http://www.todobr.com.br/>

<https://web.archive.org/web/20020401204157/http://www.todobr.com.br/>

<https://web.archive.org/web/20020401204157/http://www.todobr.com.br/>

UOL.COM.BR

<https://web.archive.org/web/20040627050910/http://busca.uol.com.br/>

<https://web.archive.org/web/20040710120204/http://mdi.uol.com.br/ciencia/historia/index.html>

<https://web.archive.org/web/19961223175947/http://uol.com.br/>

<https://web.archive.org/web/19961223175947/http://uol.com.br/>

<https://web.archive.org/web/20020802152014/http://www.uol.com.br/busca/>

<https://web.archive.org/web/20020805003343/http://radaruol.uol.com.br/radaruol.html?pg=a>

<https://web.archive.org/web/20020803052419/http://www.uol.com.br/bibliot/>

<https://web.archive.org/web/20020808015716/http://radaruol.uol.com.br/>

<https://web.archive.org/web/20040730143924/http://www.brasilnoar.com.br/>

<https://web.archive.org/web/20040803060224/http://www.multirio.rj.gov.br/historia/>

WITH IN-LINKS

<https://web.archive.org/web/20040804185846/http://mdi.uol.com.br/ciencia/historia/brasil/index.html>

REFERÊNCIAS

- ABBA, T. “Archiving Digital Narrative: Some Issues.” *Convergence: The International Journal of Research into New Media Technologies* 18.2 (2012): 121–125. Web. 21 Mar. 2014.
- ABREU, Regina. *Chicletes eu misturo com bananas? Acerca da relação entre teoria e pesquisa em memória social*. IN: In: Gondar, Jô e Dodebei, Vera. *O que é memória social?* Rio de Janeiro; Contracapa/PPGMS, 2005.
- ALLISON, Arthur et al. “Digital Identity Matters.” *Journal of the American Society for Information Science and Technology* 56.4 (2005): 364–372. Web. 21 Mar. 2014.
- AULISIO, Sandra M. et al. “Towards Brazilian Portuguese Automatic Text Simplification Systems.” *Proceeding of the eighth ACM symposium on Document engineering - DocEng '08* 24 (2008): 240.
- AMADO, Janaína e FERREIRA, Marieta de Moraes. *Usos e Abusos da História Oral*. Rio de Janeiro, FGV,1998.
- ARMS, William Y et al. “Building a Research Library for the History of the Web.” (2006): 95–102. Print.
- ARTHUR, P. “History in Motion: Digital Approaches to the Past.” (2009): n. pag. Web. 21 Mar. 2014.
- ARTIERES, Philippe . *Arquivar a própria vida. Estudos Históricos, 1998*. Disponível em: www.cpdoc.fgv.br (acessado em 15 de fevereiro de 2015).
- BARNET, Belinda. “Technical Machines and Evolution.” *CTheory* (2004): 1–19. Web. 21 Mar. 2014.
- _____. “Continuum: Journal of Media & Cultural Studies Pack-Rat or Amnesiac? Memory, the Archive and the Birth of the Internet.” July 2013 (2010): 37–41. Print.
- BAROCAS, Solon; HOOD, Sophie; ZIEWITZ, Malte. *Governing algorithms: A provocation*. 2013. Disponível em: <http://thelawsofrobitics2013.iankerr.ca/files/2013/09/48-Governing-Algorithms.pdf>. Acesso em 14/10/2015.
- BARTHEL, Ralph et al. “An Internet of Old Things as an Augmented Memory System.” *Personal and Ubiquitous Computing* 17.2 (2011): 321–333. Web. 19 Mar. 2014.
- BENJAMIN, Walter. *Magia e Técnica: Arte e Política*. Obras Escolhidas. São Paulo, Brasiliense, 1993.
- BERNERS-LEE, Tim e O'REILLY, Tim. ‘A Conversation with Tim Berners-Lee.’” *Web 2.0 Summit 09*. (2009) Disponível em: <http://www.web2summit.com/web2009/public/schedule/detail/9286> (acessado em 15 de fevereiro de 2015).

BERNERS-LEE, Tim. "Information Management: A Proposal." *Word Journal Of The International Linguistic Association* February 2 (1989): 1–10.

BERRY, David M. "The Computational Turn: Thinking about the Digital Humanities." *Culture Machine* 12 (2011): 1–22. Web. 21 Mar. 2014.

_____. "The Social Epistemologies of Software." *Social Epistemology* 26.3-4 (2012): 379–398. Web. 21 Mar. 2014.

BINGHAM, Nicola. Quality Assurance Paradigms in Web Archiving Pre and Post Legal Deposit. *Alexandria*, 2014, 25.1-2: 51-68.

BOONSTRA, BREURE, DOORN. *Past, present and future of historical information science*. Historical Social Research / Historische Sozialforschung, Vol. 29 (2004), No. 2. Disponível em: http://www.niwi.knaw.nl/en/geschiedenis/onderzoek/onderzoeksprojecten/ppf_of_his/final_report/toonplaatje (acessado em 10 setembro de 2015).

BRIGGS, Asa e BURKE, Peter. *Uma história social da mídia – de Guttenbergh à internet*. Rio de Janeiro: Jorge Zahar, 2006.

BRIN, Sergey and PAGE, Lawrence Page. "The Anatomy of a Large-Scale Hypertextual Web Search Engine." *Computer Networks* 30(1-7):107-117, 1998. 2. 1.

BROTHMAN, Brien. "Perfect Present, Perfect Gift: Finding a Place for Archival Consciousness in Social Theory." *Archival Science* 10.2 (2010): 141–189. Web. 21 Mar. 2014.

BROUWER, Joke atl all (orgs.) *Information is alive: art and theory on archiving ansretrieving data*. Rotterdam: Intituto V2, 2003.

BRÜGGER, Niels. The archived website and website philology – a new type of historical document? *Nordicom Review*, número 29 , vol. 2, 2008, pp. 155-175.

BRÜGGER, Niels. "Website History and the Website as an Object of Study." *New Media & Society* 11.1-2 (2009): 115–132. Web. 21 Mar. 2014.

BRÜGGER, Niels, and Henrik Bødker. *The Internet and Society ? The Internet and Society ? Questioning Answers and Answering*. N. p. Print.

BRÜGGER, Niels, and Niels Ole Finnemann. "The Web and Digital Humanities: Theoretical and Methodological Concerns." *Journal of Broadcasting & Electronic Media* 57.1 (2013): 66–80. Web. 21 Mar. 2014.

BURKE, Peter (org.) *A escrita da história – novas perspectivas*. São Paulo: Unesp,1992.

CASTELLS, Manuel. *A galáxia da internet – reflexões sobre a internet , os negócios e a sociedade*. Rio de Janeiro: Jorge Zahar,2003.

_____. *A sociedade em rede*. São Paulo, Paz e Terra, 2000.

CATHRO, Warwick; WEBB, Colin; WHITING, Julie. Archiving the web: The pandora archive at the national library australia. *National Library of Australia Staff Papers*, 2009.

CERTEAU, Michel de. *A Escrita da História*. Rio de Janeiro, Forense Universitária, 1982.

_____. *A invenção do cotidiano- artes do fazer*. Petrópolis: Vozes, 1994.

CHARLAB, Sergio. *Você e a Internet no Brasil*. Rio de Janeiro: Objetiva, 1995.

CHARTIER, Roger (1994). *A ordem dos livros*. Brasília, Ed. UnB.

_____. *A aventura do livro – do leitor ao navegador*. São Paulo: Ed. Unesp / Imprensa Oficial, 1999.

_____. *A História Cultural: entre práticas e representações*. Rio de Janeiro: Bertrand, 1990.

_____. *Os desafios da escrita*. São Paulo: Unesp, 2000.

CHOAY, Françoise trad. Luciano Vieira Machado. *A alegoria do patrimônio*. São Paulo: Unesp, 2001.

CHOAY, Françoise. *Le patrimoine en questions: anthologie pour un combat*. Paris: Seuil, 2009.

CHUN, Wendy Hui Kyong. The enduring ephemeral, or the future is a memory. *Critical Inquiry*, 2008, 35.1: 148-171.

_____. *Programmed visions: Software and memory*. Mit Press, 2011.

COLEMAN, E. Gabriella. “Ethnographic Approaches to Digital Media.” *Annual Review of Anthropology* 2010 : 487–505.

Cook, Terry. (1999). What is past is prologue: A history of archival ideas since 1898, and the future paradigm shift. *Archivaria*, 43, 17-63. Disponível em: <http://journals.sfu.ca/archivar/index.php/archivaria/article/view/12175/13184>. (acesso em 10/08/2014)

COSTA, Icléia Thiesen Magalhães. *Cibercultura: do caos informacional à produção de subjetividade*. Anais do XX Congresso Brasileiro de Biblioteconomia, Documentação e Informação. Ceará, junho de 2002.

CRARY, Jonathan. *Suspensions of perception: attention, spectacle and modern culture*. Cambridge(MA): MIT Press, 2001.

DANTAS, Camila Guimarães. *O passado em bits: memórias e histórias na internet*. Rio de Janeiro, Dissertação de mestrado (PPGMS/UNIRIO), março de 2008.

DANTAS, C.G. e DODEBEI, V. Dantas, Camila Guimarães, and Vera Dodebei. "Research Notes on the Emerging Concepts of Digital Heritage in Brazil." *Cultural Attitudes towards Technology and Communication 2010 Proceedings of the Seventh International Conference on Cultural Attitudes towards Technology and Communication Vancouver, Canada, 15-18 June 2010*. School of Information Technology, Murdoch University, 2010.

DARTON, Robert. "The Good Way to Do History." *New York Review of Books* 61 (2014): 52–55.

_____. *The library in the new age*. The New York Review of the Book, vol. 50, n.10, 2008. Disponível em <http://www.nybooks.com/articles/21514?email> (acessado em 10 de outubro de 2008).

_____. *The new age of the book*. The New York Review of the Book, vol. 46, n.5, 1999. Disponível em www.nybooks.com/articles/546 (acessado em 6 de dezembro de 2005).

DELALANDE, Nicolas. "PORTRAIT DE L ' HISTORIEN-NE EN CYBORG." (2011): 4–29. Print.

DELGADO, Lucilia de Almeida Neves, FERREIRA, Marieta De Moraes. "História Do Tempo Presente E Ensino de História." *História Hoje* 2 (2013): 19–34. Print.

DERRIDA, Jacques. "Archive Fever, A Freudian Impression." *Diacritics A Review Of Contemporary Criticism* 25 (2010): 9–63.

_____. *Mal de arquivo: uma impressão freudiana*. Rio de Janeiro: Relume Dumará, 1995,

DIJCK, J. *Mediated memories in the digital age*. Stanford: Stanford University Press, 2007.

DIJCK, José Van, and Thomas Poell. "Understanding Social Media Logic." 1.1 (2013): 2–14. Web. 21 Mar. 2014.

DODEBEI, Vera. *Patrimônio e memória digital*. Morpheus, Revista Eletrônica em Ciências Humanas, Ano 4 N. 8, 2006. Disponível em: <http://www.unirio.br/morpheusonline/numero08-2006/veradodebei.htm> (acessado em 10 de outubro de 2014).

_____. *Espaços míticos e imagéticos da memória social*. In: Memória e espaço. Rio de Janeiro: 7 Letras, 2000.

_____. "Digitalização Do Patrimônio E Organização Do Conhecimento." *VIII ENANCIB – Encontro Nacional de Pesquisa Em Ciência Da Informação*. N. p., 2007. Print.

_____. "Novos Meios de Memória: Livros E Leitura Na Época Dos Weblogs." *Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação* 2009. Disponível em : <https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2009v14nesp1p129> (acesso em 10/06/2015)

DONOVAN, Lori; HUKILL, Graham; PETERSON, Anna. The Web Archiving Life Cycle Model. 2013-12-12]. http://archive-it.org/static/files/archiveit_life_cycle_model.pdf, 2013.

DOSSE, François. História e Ciências Sociais. Bauru: Edusc, 2004.

ERLL, Astrid, Ansgar NÜNNING, and Sara B. YOUNG, eds. *Cultural memory studies: An international and interdisciplinary handbook*. Vol. 8. de Gruyter, 2008.

ERNST, Wolfgang. *Digital memory and the archive*. Minnesota: University of Minnesota Press, 2013

_____. "The Archive as Metaphor: From Archival Space to Archival Time." *Open* (2005): 46–53.

_____. "Does the Archive Become Metaphorical in Multi-Media Space?." *New media, old media: A history and theory reader* (2006): 105-123.

FERREIRA, Marieta de Moraes. "Demandas sociais e história do tempo presente." *Tempo presente e usos do passado*. Rio de Janeiro: Ed. Fundação Getúlio Vargas (2012).

_____. História oral: velhas questões, novos desafios. CARDOSO, *Ciro Flamarion*; VAINFAS, *Ronaldo*. *Novos domínios da história*. Rio de Janeiro: Elsevier, 2012, 169-186.

FARGE, Arlette (1989). *Le Goût de L'Archive*. Paris, Seuil, 1989.

FOGU, Claude, *Digitalizing historical consciousness*. History and Theory. Theme Issue 47 (May 2009), 1003-121.

FRAGOSO, Suely; RECUERO, Raquel; AMARAL, Adriana. *Métodos de pesquisa para internet*. 1ª Ed. Porto Alegre: Editora Sulina, 2012.

FRAGOSO, Sueli. "O Impacto Dos Buscadores Sobre O Modelo Distributivo Da World Wide Web." *Revista de Economía Política de las Tecnologías de la ...* IX (2007): n. pag. Web. 21 Mar. 2014.

GALLINI, Stefania, and NOIRET, Serge. "La Historia Digital En La Era Dossier Historia Digital." (2011): 16–37. Print.

GITELMAM, Lisa. "Media, Materiality, and the Measure of the Digital: Or, the Case of Sheet Music and the Problem of Piano Rolls." *Memory Bytes: History, Technology, and Digital Culture*. N. p., 2004. 199–217.

_____. *Always already new media: media, history and the data of culture*. Massachusetts: MIT Press, 2007.

_____. "Welcome to the Bubble Chamber: Online in the Humanities Today." *The Communication Review* 13.1 (2010): 27–36. Web. 21 Mar. 2014.

_____. *Always already new: Media, history and the data of culture*. MIT press, 2006.

_____. *Raw data is an oxymoron*. MIT Press, 2013.

GOLDMAN, E. “Search Engine Bias and the Demise of Search Engine Utopianism.” *Web Search. Information Science and Knowledge Management, Vol. 14*. N. p., 2008. 121–133.

GOMES, Daniel et al. “Search the Past with the Portuguese Web Archive.” *22nd International Conference on World Wide Web, WWW 2013*. N. p., 2013. 321–324.

_____. “Introducing the Portuguese Web Archive Initiative.” (2009): n. pag. Web. 21 Mar. 2014.

GONÇALVES, J. Reginaldo Santos. O patrimônio como categoria de pensamento. In: ABREU, Regina; CHAGAS, Mário (Orgs.) *Memória e Patrimônio. Ensaio contemporâneos*. Rio de Janeiro: DP&A, 2003. p. 21-29.

GONÇALVES, Pedro Paleio. “BRight : A Distributed System for Web Lnfomation Lndexing and Searcbing 2 . Centralized Web Lndexing and Searching.” 331–345.

GONDAR, Jô. (2005). *Quatro proposições sobre memória social*. In: Gondar, Jô e Dodebei, Vera. *O que é memória social?* Rio de Janeiro; Contracapa/PPGMS, 2005.(p.11-27).

HALAVAIS, Alex (2008) *Search Engine Society*. Cambridge: Polity.

HALBWACHS, Maurice. *A memória coletiva*. São Paulo: Editora revista do tribunais, 1990 (primeira edição 1950).

HALLAM, J. et all. “Mapping, Memory and the City: Archives, Databases and Film Historiography.” *European Journal of Cultural Studies* 14.3 (2011): 355–372. Web. 21 Mar. 2014.

HARTOG, François. *Tempos do mundo, História, Escrita da História*. In: Salgado, Manuel Luiz Guimarães (org.). *Estudos sobre a Escrita da História*. Rio de Janeiro: Sete Letras, 2006.

HASKINS, Ekaterina. “Between Archive and Participation: Public Memory in a Digital Age.” *Rhetoric Society Quarterly* 37.4 (2007): 401–422. Web. 21 Mar. 2014.

HAYLES, N. Katherine. “How We Think: Transforming Power and Digital Technologies.” *Understanding Digital Humanities*. N. p., 2012. 42–66. Print.

_____. “Hyper and Deep Attention: The Generational Divide in Cognitive Modes.” *Profession* 2007 : 187–199.

_____. “Refiguring the Posthuman.” *Comparative Literature Studies* 2004 : 311–316.

_____. “Traumas of Code.” *Critical Inquiry* 2006 : 136–157.

HEDSTROM, Margaret. “Archives , Memory , and Interfaces with the Past.” (2002): 21–43. Print.

HELMOND, Anne. “The Algorithmization of the Hyperlink.” *Computational Culture* 3 (2013): 1–37.

HEYMANN, Luciana Quillet. “Indivíduo, Memória E Resíduo Histórico: Uma Reflexão Sobre Arquivos Pessoais E O Caso Filinto Müller.” *Revista Estudos Históricos* 1997 : 41–60.

_____. De “arquivo pessoal” a “patrimônio nacional”: reflexões acerca da produção de “legados”. *Rio de Janeiro: CPDOC*, 2005, 1-18.

_____. O dever de mémoire na França contemporânea: entre memória, história, legislação e direitos. *Direitos e cidadania: memória, política e cultura*. Rio de Janeiro: FGV, 2007, 15-43.

HOCKX-YU, Helen. “The Past Issue of the Web.” ACM Web Science 2011- 3rd International Conference on Web Science Disponível em: <http://www.websci11.org/fileadmin/websci/Papers/PastIssueWeb.pdf> (acesso em 10/06/2015)

_____. *Archiving Social Media in the Context of Non-print Legal Deposit*. Paper presented at: IFLA WLIC 2014 - Lyon - Libraries, Citizens, Societies: Confluence for Knowledge in Session 107 - National Libraries. In: IFLA WLIC 2014, 16-22 August 2014, Lyon, France.

HOSKINS, Andrew. “From Collective Memory to Memory Systems.” *Memory Studies* 4 (2011): 131–133.

HUHTAMO, Erkki (ed.). *Media archaeology: Approaches, applications, and implications*. Univ of California Press, 2011.

HUTTON, Patrick. *Recent scholarship on memory and history*. The history teacher, vol. 33, n. 4 , 2000,533-548.

HUURDEMAN, Hugo C et al. “Sprint Methods for Web Archive Research.” *Proceedings of the 5th Annual ACM Web Science Conference on - WebSci '13*. N. p., 2013. 182–190.

HUYSSSEN, Andreas. *Seduzidos pela memória*. Rio de Janeiro, Aeroplano, 2000.

INSTITUTO DE PESQUISAS V2. *Capturing Unstable Media*. Holanda,2003.Disponível em: <http://www.v2.nl/> (acessado em 10 de outubro de 2015).

JAFFE, Elliot, and KIRKPATRICK, Scott. “Architecture of the Internet Archive.” *Proceedings of SYSTOR 2009: The Israeli Experimental Systems Conference on - SYSTOR '09*. N. p., 2009. 1.

JOHSON, Steven. *Cultura da interface – como o computador transforma nossa maneira de criar e comunicar*. Rio de Janeiro: Zahar, 2001.

JULIA, Dominique. BOUTIER, Jean. *Passados recompostos – campos e canteiros da História*. Rio de Janeiro: Editora FGV/Editoras UFRJ,1998.

KEEN, Andrew. *The cult of amateur – how today's internet is killing our culture*. New York : Doubleday/Currency, 2007.

KETELAAR, Eric. Tacit narratives: the meanings of archives. *Archival Science*, 2001, 1.2: 131-141.

_____. Eric. Archives in the digital age: new uses for an old science. *Archivaria*, 2001, 52: 1-25.

KLEIN, Kerwin Lee. *On the emergency of memory in historical discourse*. Representations. N. 69. Special Issue: Grounds for remembering. Winter, 2000 (p. 127-150)

KUNNY, Terry. "A Digital Dark Age? Challenges in the Preservation of Electronic Information," *International Preservation News* no. 17 (May 1998). (1997).

LAGROU, Pieter. Historiographie de guerre et historiographie du temps présent : cadres institutionnels en Europe occidentale, 1945-2000 » in *Bulletin du Comité international d'histoire de la deuxième guerre mondiale*, vol. 30-31, 1999-2000, pp. 191-215. En ligne : http://www.ihtp.cnrs.fr/equipe/Lagrou/historiographie_pl.html. (acessado em janeiro de 2007)

LE GOFF, J. Memória. *Verbete Enciclopédia Einaudi*. Campinas: Editora Unicamp, 1998.

LEJEUNE, Philippe. *Pour l'autobiographie*. Paris: Éditions du Seuil, 1996.

LEMOS, André. "Cibercultura: Alguns Pontos Para Compreender a Nossa Época." *Olhares Sobre a Cibercultura*. N. p., 2003. 11–23.

LIU, Alan. The state of the digital humanities A report and a critique. *Arts and Humanities in Higher Education*, 2012, 11.1-2: 8-41.

LÉVY, Pierre. *Abrir o espaço semântico em prol da inteligência coletiva*. RECIIS – R. Eletr. de Com. Inf. Inov. Saúde. Rio de Janeiro, v. 1, n. 1, p. 129-140, jan-jun, 2007. Disponível em: <http://www.reciis.cict.fiocruz.br/index.php/reciis/article/view/43/37> (acessado em 15 de fevereiro de 2015).

_____. *As tecnologias da inteligência – o futuro do pensamento na era da informática*. Rio de Janeiro: Ed. 34, 1993.

LOVINK, Geert (2003). "Archive Rumblings: Interview with German Media Archeologist Wolfgang Ernst." Disponível em: www.nettime.org. (acesso em 10/06/2015)

_____, Geert. *Networks without a cause: A critique of social media*. Polity Press, 2011.

SUPPIA, Alfredo Luiz; SILVA, Cícero Inácio da. Propostas e possibilidades acerca de um cinema em grid no Brasil. *Rumores-Revista de Comunicação, Linguagem e Mídias*, 2010.

LYMAN, P (2002). “Archiving the World Wide Web.” In: *Building a national strategy for digital preservation*. Washington, D.C.: Council on Library and Information Resources and Library of Congress, 2002. Disponível em: <http://www.clir.org/pubs/abstract/pub106abst.html> (acessado em 10/10/2015).

MANOFF, Marlene. “Archive and Database as Metaphor: Theorizing the Historical Record.” *portal: Libraries and the Academy* 2010 : 385–398.

_____, Marlene. “The Materiality of Digital Collections: Theoretical and Historical Perspectives.” *portal: Libraries and the Academy* 2006 : 311–325.

_____, Marlene. “Unintended Consequences: New Materialist Perspectives on Library Technologies and the Digital Record.” *portal: Libraries and the Academy* 13 (2013): 273–282.

MANOVICH, Lev (1999). “Database as Symbolic Form.” *Convergence: The International Journal of Research into New Media Technologies* 5.2 (1999): 80–99.

_____, *Software takes command*. (2008) Disponível em: www.softwarestudies.com/ (acessado em 10 de julho de 2014).

_____, *The language of new media*.(2001) Disponível em: www.softwarestudies.com/ (acessado em 10 de julho de 2010).

MARCUSCHI, LA, XAVIER,AC (orgs). *Hipertexto e gêneros digitais*.Rio de Janeiro, Lucerna, 2005.

MARTY, Paul F. “An Introduction to Digital Convergence: Libraries, Archives, and Museums in the Information Age.” *Archival Science* 8 (2008): 247–250.

MATTELART, Armand. *História da sociedade da informação*. São Paulo: Loyla, 2001.

MAYER-SCHÖNBERGER, Viktor. *Useful Void: The Art of Forgetting in the Age of Ubiquitous Computing*. Faculty Research Working Papers Series, John F. Kennedy School of Government - Harvard University, 2007.

MENKMAN, Rosa. *The glitch moment (um)*. Amsterdam: Institute of Network Cultures, 2011.

MEDIAMATIC. WEBZINE. 1994. Disponível em: <http://www.mediamatic.net/245164/nl/dit-is-mediamatic>. Acesso em 14 de outubro de 2015.

MERLIN, D. *Is a pictures really worth 1000 words?* History and Theory. N. 43, Outubro 2004, pp 379-385.

MINUTI, Rolando. *Internet et le métier d'historien – réflexions sur les incertitudes d'une mutation*. Paris: Presses Universitaires de France, 2002.

CARELLI, Ana Esmeralda; MONTEIRO, Silvana Drumond. *Ciberespaço, memória e esquecimento*. 2007.

NEGROPONTE, Nicholas. *A vida digital*. São Paulo: Companhia das Letras, 1995.

NORA, Pierre. *Entre a memória e a história: a problemática dos lugares*. Projeto História, PUC-SP, dez.1993.

NOVICK, P. *That noble dream. The objectivity question and the american historical profession*. Cambridge: Cambridge University Press, 1988.

O'REILLY, Tim. *What is web 2.0?* (2005). Disponível em : <http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html> (acessado em 10/06/2015)

OLICK, Jeffrey K.; ROBBINS, Joyce. Social memory studies: From "collective memory" to the historical sociology of mnemonic practices. *Annual Review of sociology*, 1998, 105-140.

_____, OLICK, Jeffrey K. From collective memory to the sociology of mnemonic practices and products. *Cultural memory studies: an international and interdisciplinary handbook*, 2008, 8: 151.

OMODEI, Monica e Gordon Mohr. "Internet Content as Research Data." March (2012): n. pag. Print.

Osborne, T. "The Ordinarity of the Archive." *History of the Human Sciences* 1999: 51-64.

PARENTE, André (org.) *Tramas da rede – novas dimensões filosóficas, estéticas e políticas da comunicação*. Porto Alegre: Sulina, 2004.

PARIKKA, Jussi e Tony D Sampson. "On Anomalous Objects of Digital Culture." *The Spam Book: On Viruses, Porn, and Other Anomalies from the Dark Side of Digital Culture*. N. p., 2009.

_____, Jussi. *What is media archaeology*. John Wiley & Sons, 2013.

POLLACK, Michael. *Memória, Esquecimento, Silêncio*. Estudos Históricos, Rio de Janeiro, vol.2, nº 3, 1989.

PORTELLI, Alessandro. *Field notes from Harlam County*. Disponível em: <http://www.albany.edu/jmmh/http://www.albany.edu/jmmh/>)

REIFF, Janice I. *Structuring the past: the use of the computer in History*. Washington, D.C: American Historical Association, 1991.

REVEL, Jacques. *História e ciências sociais: uma confrontação instável*. In: JULIA, Dominique. BOUTIER, Jean. *Passados recompostos – campos e canteiros da História*. Rio de Janeiro: Editora FGV/Editoras UFRJ, 1998.

RICOUER, Paul. *A memória, a história, o esquecimento*. Campinas: Editora Unicamp, 2007.

RICOUEUR, Paul. Entrevista. *Historia, Antropologia y Fuentes Orales*. N. 30, 2003.

RIEDER, B., and G. Sire. "Conflicts of Interest and Incentives to Bias: A Microeconomic Critique of Google's Tangled Position on the Web." *New Media & Society* 16.2 (2013): 195–211.

ROGERS, Richard (2010). "Internet Research: The Question of Method," *Journal of Information Technology and Politics*, 7, 2/3, 2010, 241-260.

_____, Richard. (2013a) "Internet Research: The Question of Method," in: Christine Hine (ed.), *Virtual Research Methods*. London: Sage, 2013, 337-364.

_____, Richard. (2013b) *Digital Methods*, Cambridge, MA: MIT Press, 2013.

RONDINELLI, Rosely Curi. *Gerenciamento eletrônico de documentos eletrônicos*. Rio de Janeiro, FGV, 2005.

ROSSITER, Ned; LOVINK, Geert. *My Creativity Reader: A Critique of Creative Industries*. Amsterdam: Institute of Network Cultures, 2007.

ROUSSO, Henry. "A memória não é mais o que era". IN: FERREIRA, Marieta de Moraes & AMADO, Janaína. *Usos e Abusos da História Oral*. RJ: FGV, 2002.

_____. "O arquivo ou indício de uma falta". *Estudos Históricos*, Rio de Janeiro, n.17, 1996.

ROZENZWEIG, Roy. *Scarcity or Abundance? Preserving the Past in a Digital Era*. *American Historical Review* 108, 3 (June 2003): 735-762 Disponível em: <http://chnm.gmu.edu/resources/essays/d/6> (acessado em 12 de agosto de 2008).

SÁ, A. F. DE A. *Admirável campo novo: o profissional de história e a Internet*. Rio de Janeiro: Revista Eletrônica Boletim do TEMPO, Ano 3, n. 07, Rio, 2008. [Disponível em http://www.tempopresente.org/index.php?option=com_content&task=view&id=3620&Itemid=124. Acessado em outubro de 2008]

SALGADO, Manuel Luiz Guimarães (org.). *Estudos sobre a Escrita da História*. Rio de Janeiro: Sete Letras, 2006.

SANDERSON, Robert. "Global Web Archive Integration with Memento." *Proceedings of the 12th ACM/IEEE-CS Joint Conference on Digital Libraries*. N. p., 2012. 379–380.

SANTOS, Myriam Sepúlveda. *Memória coletiva e teoria social*. São Paulo: Anablume, 2003.

SARLO, Beatriz. *Tempo Passado – cultura da memória e guinada subjetiva*. São Paulo, Cia das Letras; Belo Horizonte, UFMG, 2007.

SCHITTINE, Denise. *Blog: comunicação e escrita íntima na internet*. Rio de Janeiro: Civilização Brasileira, 2004.

SCHWARTZ, Joan M., and Terry Cook. "Archives, Records, and Power: The Making of Modern Memory." *Archival Science* 2.1-2 (2002): 1–19.

SNICKARS, Pelle. "The archival cloud". In: Snickars, P e Vonderau, P (orgs.). *The YouTube reader*. Stockholm: National Library of Sweden, 2009, pp. 292–313.

SOLBERG, Janine. "Googling the Archive: Digital Tools and the Practice of History." *Advances in the History of Rhetoric* 15.1 (2012): 53–76.

STALEY, David J. *Computers, vizualization and History – how new technology will transform our understanding of the past*. New York: M.E Sharpe, 2003.

STEPHENS, M., and R. Neill. "'You Do Know It's Only Got One Bathroom': Biography, Wine and Small Stories." *Memory Studies* 4.2 (2010): 221–233. Web. 21 Mar. 2014.

STURKEN, Marita. *Tangled memories: The Vietnam War, the AIDS epidemic, and the politics of remembering*. Univ of California Press, 1997.

_____. "Narratives of recovery: Repressed memory as cultural memory." *Acts of memory: Cultural recall in the present* (1999): 231-48.

_____. "The aesthetics of absence: Rebuilding ground zero." *American Ethnologist* 31.3 (2004): 311-325.

TEDESCO, João Carlos. *Nas cercanias da memória – temporalidade, experiência e narração*. Passo Fundo, UPF; Caxias do Sul, EDUCS, 2004.

THELWALL, Mike; VAUGHAN, Liwen. A fair history of the Web? Examining country balance in the Internet Archive. *Library & information science research*, 2004, 26.2: 162-176.

TODOROV, Tzvetan. *Memória do mal, tentação do bem – indagações sobre o século XX*. São Paulo, Arx, 2002.

TOYODA, Masashi; KITSUREGAWA, Masaru. "The History of Web Archiving." *Proceedings of the IEEE*. Vol. 100. N. p., 2012. 1441–1443.

TURKEL, William J. "Rethinking History: The Journal of Theory and Practice Intervention : Hacking History, from Analogue to Digital and Back Again." August 2012 (2011): 37–41. Print.

UKOLN. *Collecting and preserving the World Wide Web*. Londres, 2003. Disponível em <http://www.ukoln.ac.uk/>

UNESCO. *Basic Texts on the Information Society*. 2003. Disponível em: http://portal.unesco.org/en/ev.phpURL_ID=21647&URL_DO=DO_TOPIC&URL_SECTION=201.html (acessado em 10 de setembro de 2008).

VAIDHYANATHANOU, Siva (2011). *The Googlization of Everything (And Why We Should Worry)*, Berkley: University of California Press, 2011. Winterson, Jeannette.

VAN DIJCK, J. “Flickr and the Culture of Connectivity: Sharing Views, Experiences, Memories.” *Memory Studies* 4.4 (2010): 401–415.

_____. “Search Engines and the Production of Academic Knowledge.” *International Journal of Cultural Studies* 13.6 (2010): 574–592.

VAN HOUSE, Nancy; CHURCHILL, Elizabeth F. Technologies of memory: Key issues and critical perspectives. *Memory Studies*, 2008, 1.3: 295-310.

VASCONCELOS, Nelson. *Tem alguém perdido aí?* Informática&Etc. Jornal O Globo. 14/04/1997

VELICU, Adrian. “Cultural Memory between the National and the Transnational.” *Journal of AESTHETICS & CULTURE* 3 (2011): 1–4. Web. 21 Mar. 2014.

VIEGAS, Fernanda. *Revealing Individual and Collective Pasts: Visualizations of Online Social Archives*. Tese de Doutorado, MIT, 2005. Disponível em: http://smg.media.mit.edu/papers/Viegas/viegas_thesis.pdf (acessado em 05 de setembro de 2007)

ZIEROLD, Martin. Memory and media cultures. In:ERLL, Astrid; NÜNNING, Ansgar; YOUNG, Sara B. (ed.). *Cultural memory studies: An international and interdisciplinary handbook*. de Gruyter, 2008,pp399-408.

ZIMMER, Michael. The externalities of search 2.0: The emerging privacy threats when the drive for the perfect search engine meets Web 2.0. *First Monday*, 2008, 13.3.

ZINS, Chaim. “Redefining Information Science: From ‘information Science’ to ‘knowledge Science.’” *Journal of Documentation* 62.4 (2006): 447–461. Web. 21 Mar. 2014.

ZIMMER, Michael. *The gaze of the perfect search engine: Google as an infrastructure of dataveillance*. Springer Berlin Heidelberg, 2008.